

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

The Principle of Fairness and Natural Duties

Permalink

<https://escholarship.org/uc/item/6fw181w3>

Author

Finley, Aaron

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Principle of Fairness and Natural Duties

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of
Philosophy

in

Philosophy

by

Aaron Finley

Committee in Charge:

Professor Richard Arneson, Co-Chair
Professor Saba Bazargan-Forward, Co-Chair
Professor Sean Ingham
Professor Andy Lamey
Professor Samuel Rickless

2021

Copyright

Aaron Finley, 2021

All rights reserved.

The dissertation of Aaron Finley is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

EPIGRAPH

There are also many positive acts for the benefit of others, which [one] may rightfully be compelled to perform; such as...to bear his fair share in the common defence, or in any other joint work necessary to the interest of the society of which he enjoys the protection.

John Stuart Mill

TABLE OF CONTENTS

Dissertation Approval Page	iii
Epigraph.....	iv
Table of Contents.....	v
Acknowledgments.....	vii
Vita.....	viii
Abstract of the Dissertation	ix
Chapter 1, Opportunity Costs and the Value of Benefits in the Principle of Fairness	1
Introduction.....	1
1.1 The Principle of Fairness	3
1.1.1 Public Goods and Ex Ante Price Negotiation.....	5
1.1.2 Mutual Advantage and Proportionality.....	11
1.2 Accepting Benefits Willingly and Knowingly.....	13
1.2.1 Adopting the Knowledge Condition	16
1.2.2 Is Non-Culpable Ignorance Blissful Ignorance?.....	19
1.2.3 Simmons’s Preference-Based Theory of Value.....	23
1.2.4 Opportunity Costs: Free-Riders and Efficiency.....	26
1.3 Conclusion	30
Chapter 2, Fair Play and Artificial Agents.....	33
Introduction.....	33
2.1 Mediterranean Migrants and International Law	34
2.2 Lefkowitz on the Mental States of States	38
2.3 Perfectionism as State Well-Being	44
2.4 Alignment of Individual and State Well-Being	52
2.5 Causal Efficacy of State Obligations	58
Chapter 3, Fairness Without Cooperation.....	62
Introduction.....	62
3.1 Achilles Among the Farmers.....	66
3.2 Fair Play as a Transactional Relationship.....	70
3.3 Transactional Relationships as Legitimate Expectations of Reciprocation.....	74
3.4 Overgeneralization.....	77
3.5 Conclusion	81
Chapter 4, Slack-Taking and Burden-Dumping: Fair Cost-Sharing in Duties to Rescue	86
Introduction.....	86
4.1 The No-Burden-Dumping Intuition	89
4.2 Collective Burden-Dumping.....	93

4.3 The Particularity Problem	100
4.4 Omission and Collective Action	103
4.5 Further Applications	107
Chapter 5, Multiple Principles of Political Obligation	114
Introduction.....	114
5.1 Multiple-Principle (MP) Theory.....	118
5.2 The Natural Duty of Justice	122
5.2.1 Waldron.....	123
5.2.2 Wellman	128
5.2.3 Justice and Specificity	133
5.3 The Principle of Fairness and Generality.....	139
5.4 The Principle of Fairness and Completeness.....	144
5.5 Reciprocity.....	147

ACKNOWLEDGMENTS

I would like to acknowledge Professor Richard Arneson for his work as co-chair of my committee. Our regular conversations, and his detailed and wide-ranging comments on my work have been invaluable for the development of my work, and my development as a philosopher.

I would like to acknowledge Professor Saba Bazargan-Forward for his work as co-chair of my committee. It is through his guidance and instruction that I have been able to persevere to the end.

I would like to acknowledge Professor Andy Lamey for his work on my committee. Not only has he provided valuable feedback on multiple drafts, he has helped me cultivate my skills as an instructor and my love for teaching.

Chapter 4, in full, has been submitted for publication of the material as it may appear in the *Journal of Ethics and Social Philosophy*, 2021, Finley, Aaron, Elsevier, 2021. The dissertation author was the primary author of this paper.

Chapters 1, 2, 3, and 5, in full or in part, are currently being prepared for publication of the material. Finley, Aaron. The dissertation author was the primary author of this material.

VITA

2013 Bachelor of Science, University of Wisconsin Madison
2013-2020 Teaching Assistant, University of California San Diego
2019-2021 Primary Instructor, University of California San Diego
2021 Doctor of Philosophy, University of California San Diego

FIELDS OF STUDY

Major Field: Philosophy

Studies in moral and political philosophy
Professors Richard Arneson and Saba Bazargan-Forward

ABSTRACT OF THE DISSERTATION

The Principle of Fairness and Natural Duties

by

Aaron Finley

Doctor of Philosophy in Philosophy

University of California San Diego, 2021

Professor Richard Arneson, Co-Chair
Professor Saba Bazargan-Forward, Co-Chair

The Principle of Fairness and Natural Duties defends the Principle of Fairness (PoF) as a widely applicable moral principle that can ground obligations to obey the law. The first three chapters develop an interpretation of the PoF by responding to criticisms from A. John Simmons. Simmons argues that, for the PoF to generate obligations, beneficiaries of cooperative public-goods schemes must voluntarily accept the benefits they receive. I argue against Simmons's voluntarism by showing that, even on his view, mere receipt of benefits that are worth their cost

can trigger fairness obligations. In a similar vein, Simmons argues that the PoF requires a type of cooperation that we do not see in political communities. Thus, even if his voluntarism argument fails, Simmons can conclude that the PoF cannot ground political obligations. In response, I argue that the cooperation condition can be removed from the PoF entirely. Simmons's arguments defend the broader claim that the PoF applies to few individuals and generates few obligations. In pushing back against this conclusion, I argue that the PoF applies not only to natural agents, such as humans, but also to some artificial agents such as governments and corporations.

While the PoF is a robust moral principle, it may not be able to satisfactorily ground political obligation. The last two chapters of this dissertation, therefore, explore the possibility of combining the PoF with natural duties. This combination has been attempted several times in the literature on political obligation. Most often, the PoF is combined with some version of the natural duty to rescue. I argue that this duty can be combined with the PoF, but that the combination cannot ground political obligation. Rather, they ground an obligation to perform one's fair share of global rescues, though each individual has some discretion in choosing which rescue efforts to contribute to. To ground political obligation, the PoF must be combined with the natural duty of justice. I argue that the two principles are mutually reinforcing, answering objections that each, alone, cannot overcome. This combination leads to a novel theory of political obligation.

Chapter 1

Opportunity Costs and the Value of Benefits in the Principle of Fairness

Introduction

Since its initial formulation by H.L.A. Hart in 1955, the Principle of Fairness (which I will often refer to simply as ‘the principle’) has received extensive scholarly attention. In order to avoid recounting familiar debates, I will use a formulation of the principle recently defended by Richard Arneson as the basis of my discussion. He states the principle as follows:

When a number of persons engage in a just, mutually advantageous cooperative venture according to rules and thus restrain their liberty in ways necessary to generate non-excludable and non-optional benefits for all, those who have submitted to these restrictions have a right to similar acquiescence on the part of those who have benefited from their submission, at least when failure to acquiesce in this way will impose costs on the cooperators. If the goods provided are non-excludable but optional, those who do not exercise the option of taking the goods provided do not acquire obligations under this principle. (Arneson 2013, 151)

The central intuition the principle is meant to capture is that it is wrong to be a free-rider—one who benefits from the cooperative sacrifices of others without bearing her fair share of the burdens. The difficulty is to embody this intuition in a precise and defensible principle that coheres sufficiently well with our other moral intuitions and considered beliefs, and that can adequately block objections that have been advanced against it. Arneson’s formulation of the principle already takes us far along this path, but not, in my view, far enough. In particular, we can uncover new insights into the debate between voluntarists (most notably A. John Simmons) and non-voluntarists about the adequacy of the principle of fairness by examining Simmons’s claim that a beneficiary of a cooperative scheme must accept its benefits before obligations of fair play can arise for her.

In his own work, Simmons develops two main lines of argumentation against the principle of fairness.¹ First, he argues that mere receipt of benefits cannot generate obligations under the principle of fairness. Instead, obligations can only arise once the benefits one receives are *accepted*. Second, leveraging the principle's cooperation requirement, Simmons defends a very thick conception of cooperation. He maintains that once we have taken this conception on board, we will see that modern political societies do not constitute genuinely cooperative schemes.

In this chapter I will focus exclusively on Simmons's first argument—that one can have obligations of fair play only after accepting the benefits of a cooperative scheme. Before taking up my arguments against Simmons, I will need to explain how I understand the principle itself. So, in section one I lay out what I take to be the moral underpinnings of the principle of fairness that are relevant for my purposes in this chapter. In section two I take up Simmons's notion of acceptance of benefits—a notion that involves a problematic analysis of opportunity costs. Simmons apparently argues that the benefits of a cooperative scheme cannot be worth their cost unless they are optimally provided according to the preferences of recipients. Everyone agrees that the principle only generates obligations when the benefits of a scheme outweigh its costs, but surely a beneficiary need not receive those benefits in the way she most prefers for this requirement to be satisfied. Once we see this, we will see that it is much easier than Simmons supposes for members of modern political societies to accept the benefits that flow from cooperative schemes. And, of course, the one who accepts benefits incurs obligations of fairness.

To the extent that those involved in debates about the principle of fairness are concerned with voluntarism and Simmons's arguments against the principle's adequacy, the conclusions I

¹ Simmons 1979, ch. 5; 2001, ch. 2. My overview of Simmons's arguments here closely follows his own presentation of them in the introduction of chapter two of *Justification and Legitimacy* (Simmons 2001, 27-29).

draw should push us to think not in terms of voluntarism versus non-voluntarism, but in terms of differing accounts of preference satisfaction. It turns out that the real question at issue is not, ‘Can obligations of fair play be acquired non-voluntarily?’ but, ‘How do benefits come to be worth their cost?’ Simmons urges us to take opportunity costs seriously when we are tallying the costs of a cooperative scheme, and I believe he is right to do so. What I hope to show, however, is that by tallying carefully, we can allay some of Simmons’s skepticism about the adequacy of the principle, and give a more fine-grained account of what different sorts of individuals ought to contribute to a scheme in virtue of the benefits they receive from it.

1. The Principle of Fairness

Not all unfairness is bad. If I play chess with my very young niece, it is okay for her to change the rules often and arbitrarily thereby guaranteeing that she wins. Even though I have submitted to the rules of a cooperative scheme (by following the game’s rules), there’s nothing wrong with her failure to comply. However, unfair behavior from a finalist at the World Chess Championship certainly would not be acceptable. The principle of fairness, then, is intended to help us identify a type of unacceptable unfairness that is especially relevant for political societies. Roughly speaking, the principle’s claim is that one should not free-ride on the cooperative efforts of others, at least when one’s failure to cooperate increases the cost of cooperation for others in the scheme.² For instance, if three of my friends and I are maintaining a trench that supplies us with fresh running water by dredging it periodically, and I stop

² However, obligations of fair play are not generated when a fair-play-governed scheme merely redresses harms. If, for instance, city A pollutes both its own air and the air of neighboring city B, residents of city B should not be obligated to contribute to a scheme instituted in city A to reduce its pollution. Even though those in city B benefit, even after factoring in the costs of participation, the scheme serves them merely by redressing a wrong committed against them. They are not obligated to contribute because those in city A only count as making reparations.

contributing, my three friends will now have to shoulder my portion of the dredging responsibilities. Because I fail to cooperate, extra costs are wrongfully imposed on them.

The previous example is intuitive enough, but the principle of fairness, as I will defend it, covers less intuitive cases as well. Suppose I do not help create the trench scheme. Instead, my three friends institute and begin maintaining it while I am out of town. The version of the principle given above says that when I return from my trip and begin benefitting from the scheme, I am obligated to contribute. If I do not bear a fair share of the scheme's burdens, I wrong my friends even though I have done nothing to consent to the scheme, and have no history of contributing to it. So the question is, why I am obligated to the scheme in the first place? Why am I governed by its rules? To get a sense of why free-riding is wrong even in the second trench-maintenance example, we should look at the kinds of goods the principle covers, and the meaning of the terms 'mutually advantageous' and 'benefits for all' as they are used in the principle.

I should also note at the outset what I mean by 'scheme'. Perhaps surprisingly, the term is rarely defined in the literature and there has been very little discussion of how schemes should be individuated.³ While I do not take up a discussion of scheme-individuation here, I borrow a definition of 'scheme' from Simmons. As I read him, a scheme occurs when "others make deliberate sacrifices in support of mutually beneficial goals, while *relying* on others to do the same—and while having reasonable grounds for such reliance. . . [on the basis of] sharing the same goals and understandings" (Simmons 2001, 40, original emphasis). I should note that Simmons may not intend this passage as a definition of 'scheme'. The quotation comes from a discussion of cooperation (a topic I take up in Chapter 3), and, together with the surrounding

³ But see Klosko (2019) chapter 4 for some suggestions on scheme-individuation.

material, it aims to characterize cooperative schemes. Still, a characterization of cooperative schemes must include a characterization of schemes in general. The passage I have quoted is the one that gives the general characterization. Whatever we should say exegetically, however, the definition of a scheme offered above will work well for our purposes in this and subsequent chapters.

1.1 Public Goods and Ex Ante Price Negotiation

The principle of fairness is designed to cover the provision of public goods, and in this section I will highlight some of the features of these goods that make them morally and transactionally special. As I will use the term, a public good is any good that is both non-excludable and non-optional.⁴ If a good is non-excludable, then if the good is available to one person for consumption, it is not feasible to exclude others from consuming it as well. Next, if a good is non-optional, then within the region where it is provided, if anyone consumes any of the good, everyone must consume at least some of the good. A paradigm example of a good that exhibits both these features is national security. Once national security is provided to anyone within a geographic region, it is not feasible to exclude anyone in that region from receiving it. And once anyone is consuming the good, everyone must consume it.⁵ For expositional ease, I will often refer to goods that are non-excludable and non-optional simply as public goods.

Public goods are transactionally special because it is often impossible (or at least prohibitively difficult) to distribute them through voluntary transactions where the parties to the transaction agree on a price for the good *ex ante*. This is why the principle of fairness claims that

⁴ While these are the only features of public goods my arguments depend upon, for the most part I appeal to goods that are also non-rivalrous. A good is non-rivalrous if one person's consumption does not limit others' ability to consume it as well (Ostrom 2010, 642).

⁵ For a more detailed discussion of these kinds of goods, see Arneson 2013, 137,147; Ostrom 2010, 644-45.

unsolicited receipt of public goods can generate obligations in beneficiaries to reciprocate. To substantiate this claim a friend of fair play must show that the reasons that require one to engage in *ex ante* price negotiations in the case of private goods do not apply to public goods in important cases. The first question, then, is why must we engage in price negotiations in the case of private goods? There are several good answers.

George Klosko has argued that because of the importance of liberty within the liberal tradition, “there is a strong presumption that individuals should decide for themselves if they will be forced to make sacrifices or have their liberty curtailed” (Klosko 2004, 36). Thus, in the case of private goods or schemes that provide excludable goods, the burdens of participating in the scheme should not be imposed on an individual unless she voluntarily accepts the burdens and benefits of the scheme together. So out of respect for the liberty of others, one may not shower excludable goods on them expecting payment in return. In the case of public goods, however, such a constraint would almost always require non-provision since it would only take one holdout to undermine the scheme for everyone. On a practical level, non-provision would be catastrophic in the case of goods like national security. More than this, our respect for the liberty of others requires us to adopt a principle like the principle of fairness to govern the provision of public goods. While we must not be allowed to make arbitrary impositions on others, we also must not be allowed to impose arbitrary restrictions on others. A few should not be able to trample down the liberty of the rest to organize and maintain cooperative schemes. The principle of fairness aims to adjudicate this balance of liberties by defining the circumstances in which receipt of a public good generates an obligation to reciprocate, and when it does not.

Arneson presents us with a further reason for maintaining that the principle should not cover excludable goods: if possible, we should avoid situations in which holdout bargaining is

possible. Holdout bargaining occurs when an individual refuses to pay for a good (or insists on an unreasonably low price) because she has already received it. If she does this because she is psychologically unable to accurately determine the price she would have paid if bargaining had taken place *ex ante*, the problem is merely practical.⁶ Alternatively, she may holdout because of a pernicious desire to take advantage of the productive sacrifices of others by receiving the benefits they produce without paying a fair price. In the latter case, she is moved by what Arneson calls a *free rider motive* (Arneson 2013, 137). Agreeing on a price before the benefit is conferred safeguards the good-provision process against both kinds of holdout bargaining. But in the case of ongoing public-goods schemes, a price cannot be agreed upon ahead of time. It would not be feasible to attempt to negotiate a price for national security with those who are born into an ongoing national security providing scheme.

A third reason for believing that having private goods showered upon one does not generate obligations to reciprocate is offered by Robert Goodin. He argues that, even if I am indifferent between two bundles of goods (e.g. five dollars and no shoeshine versus a shoeshine and no five dollars), you may not impose one bundle on me when I would otherwise have had the opportunity to choose between the two. You may not, Goodin argues, forcibly move me along my indifference curve. Goodin cites two major obligations in support of his claim: you are obliged to respect the narrative unity of my life,⁷ and you are obligated to respect my autonomy, both of which you may violate by forcibly moving me along my indifference curve (Goodin 1989, 68-69). Moreover, these obligations seem to militate even against forcibly moving me to a

⁶ Given that the other reasons for requiring *ex ante* price bargaining are moral reasons, this reason may seem superfluous. There may be some sense in which this is right, but the practical worry is still worth mentioning; if epistemic difficulties of this sort are likely to result in serious problems in real-world bargaining situations, the importance of avoiding the harms of no provision of key public goods would be substantial.

⁷ Cf. Rudd 2007

higher indifference curve since, given the option, I may have chosen to move myself to a different location along it, or foregone the opportunity to move up altogether. But again, since one cannot feasibly opt out of a cooperative scheme, and since it would be unfair to respect narrative unity and autonomy only in the lives of those who wish to opt out of a cooperative scheme, a principle like the principle of fairness is needed to govern the provision of public goods.

Finally, since it is clear that obligations can only be incurred once benefits have actually been received, no one is obligated to contribute to a scheme that provides an optional good unless she opts to take it. If the good is deliberately taken, it is clear that the taker has bound herself to the provision scheme.⁸ So, if there is any kind of good that requires special attention, it is public goods, rather than goods that are merely non-excludable or non-optional.

We now have good reason to think that public goods are a special case in need of special treatment, at least when good-providers cannot feasibly bargain with recipients prior to the institution of the provision scheme. In such cases, benefactors cannot give beneficiaries the opportunity to forego the benefit the scheme produces. (Though of course, if *ex ante* price negotiation *is* possible, it must be pursued for the reasons listed above, even though the good in question is public.) If, however, negotiations are foregone when they might feasibly have been undertaken, any public goods the scheme produces will not give rise to obligations under the principle of fairness. Such goods might reasonably be considered a gift until negotiations are pursued with respect to future provision.⁹ This is quite different from the case in which a scheme

⁸ For a more detailed discussion, see Simmons 1979, 128-31.

⁹ Questions then arise concerning what to say about cases in which negotiations were initially feasible but were not pursued, after which circumstances change such that negotiations about future provision become unfeasible. My own intuition is to say that the principle of fairness would then begin to apply.

is already up and running and new people are born into it, immigrate, etc. Since it is not feasible to negotiate with these individuals ahead of time, there cannot be an obligation to do so.

Even when initiating a scheme, negotiations may not be feasible. A defense scheme, for instance, may need to be enacted very quickly, or the group of beneficiaries may be too large or disorganized to meaningfully participate in negotiations. In cases like these, if the conditions of the principle of fairness are satisfied, how much are beneficiaries obligated to contribute? The answer is straightforward: beneficiaries are required to contribute either the amount they would have contributed if negotiations had been possible, or their fair share of the costs of the scheme, whichever is less.

As a toy example, suppose there are ten beneficiaries of a scheme that costs \$100, and that a fair contribution from each beneficiary would be \$10. Surely it would be unfair to demand of someone that she contribute \$50 just because she would have agreed to pay that much had *ex ante* negotiations been possible. Alternatively, if someone could use morally arbitrary (or morally objectionable) features of her bargaining position to pay only \$1 when \$10 is a fair share, and the good she receives is worth more than \$10 to her, it would be unfair for her to pay less than \$10. While I will not attempt to defend a full list of these features, it should include items such as objectionable (e.g. racist) preferences, information asymmetries, and coercive power relations.

I do not appeal directly to what a good is worth to a recipient because her preferences cannot be fully determined until she undertakes the process of negotiating. The reason for this is partly epistemic. Often, we do not know what we are willing to pay for something until we are presented with the option to. More importantly, however, I cannot fully determine my preferences until I know something about yours, at least where scarce resources are involved. As

Dworkin notes, the price of a good or service, negotiated in a fair market, informs each person of “the true opportunity cost to others of his acquiring it” (Dworkin 2011, 357). My preferences ought to take into account the impact their satisfaction will have on others, which requires negotiation or some other means of effective communication.

Returning to public goods as such, the example of national defense should illustrate that a number of important public goods cannot reasonably be foregone. Thus, we need a way to determine how the provision of such goods should be handled. One could maintain that mere receipt of benefits—even in the form of public goods—does not give rise to obligations. This claim will be explored in more detail in section two, but for now I will only attempt to provisionally motivate the intuition that mere receipt of public goods can generate obligations. The intuition is, unsurprisingly, an intuition of fairness. Public-good-providers (including beneficiaries who already participate in the scheme) submit to rules and restrain their liberty in ways necessary to produce and enjoy the scheme’s benefits, so it is unfair for others to simply receive those benefits without contributing as well. Moral equals should be treated equally, so failing to contribute counts as treating participants unequally in an unfair way.¹⁰

I must make one final comment before moving on to an explanation of the last two terms I will cover in section one. Arneson, in his presentation of the principle of fairness, notes that there will usually be a range of schemes that could produce the desired good, some of which are better than others. However, the particular scheme that is implemented need not be the best, strictly speaking, for the restrictions of liberties it imposes to count as necessary for the generation of whatever good it produces. Rather, the scheme is good enough so long as it is not “significantly inferior” to other available schemes (Arneson 2013, 135-36).

¹⁰ Klosko appeals to this line reasoning in his chapter on the principle in, Klosko 2004, 34.

Here, I am understanding a scheme to be significantly inferior to another if it provides the good in question as a public good when an otherwise comparable scheme could have provided it as an excludable or optional good. Schemes that unnecessarily provide goods as public are counted inferior because they unnecessarily restrict the autonomy of beneficiaries. Thus, even when *ex ante* price negotiation is not feasible, if the good could just as well have been provided excludably or optionally, the principle of fairness will not bind. Additionally, schemes that provide a public good inefficiently relative to alternative available schemes generally will not generate obligations. In this case, obligations do not arise because the good provided will not be worth its cost once opportunity costs are taken into account. I will defend this claim in detail in section two.

1.2 Mutual Advantage and Proportionality

Let us now turn to the terms ‘mutually advantageous’ and ‘benefits for all’ to clarify what is meant by each. To say that a scheme must be mutually advantageous is to say that, for each participant, the benefits produced by the scheme cannot be outweighed by its costs.¹¹ Thus, if we all participate in a scheme, we must all be advantaged together. However, we need not all be advantaged in the same way or bear the same kind of costs. Suppose, for instance, that I have a ship, but not the expertise to keep it in good repair, and you have goods to sell, but no means of shipping them. (Suppose also that I value the repairs I need for my ship exactly as much as you value the profits you would get from selling your goods.) If you only repair my ship, only I am made better off, but if I only transport your goods (after finding some other way to repair my

¹¹ This condition assumes the cost-levels that apply when all (or nearly all) who benefit contribute to the scheme once it is off the ground. However, if the scheme has persistent free rider problems that prevent the scheme from being mutually advantageous, the scheme is a failure and the principle of fairness does not bind.

ship), only you are made better off. But if we combine the two schemes, we are advantaged mutually even though we are advantaged differently. In the same way, a scheme that produces public goods to which the principle of fairness might apply could be composed of several sub-schemes. And if the sub-schemes do not impose burdens on the same people they benefit, they will have to be bundled in a way that ensures the mutual advantage condition is met.

While a cooperative scheme does not produce obligations unless it produces mutual advantages, one may voluntarily contribute to a scheme that is not worth its cost so long as it is not unjust or otherwise intrinsically impermissible. Similarly, the principle of fairness does not tell us which schemes we are obligated to institute. There may be cooperative schemes we are obligated to put in place, but we will not violate the principle of fairness until we fail to contribute to a scheme that is up and running (Arneson 2013, 135). It is only then that failure to participate counts as free riding on the productive efforts of others.¹²

Finally, the phrase ‘benefits for all’ means that the benefit produced by the scheme must actually be a benefit before fairness obligations arise; receipt of detriments does not obligate. Then, as noted above, the benefit must continue to be a good once the costs of the scheme are factored in. If, after some time, the scheme begins producing goods that are not a net benefit to a given recipient, that individual does not have any obligation under the principle of fairness to submit to the rules of the scheme.

One last clarification that is common in the literature is that the public-goods scheme must distribute burdens in proportion to benefits. For instance, if my country’s national security scheme makes me only half as safe as everyone else, I should contribute half what everyone else

¹² Notice that, here, I do not make reference to a free rider *motive*. The idea is that, under the circumstances described, those who fail to comply simply are free-riders regardless of their motives for not submitting. When the other conditions of the principle are satisfied, all the free rider motive does is signal that the free rider thinks the benefits of the scheme are worth their cost which means that the principle of fairness applies.

does.¹³ Under all these conditions—the good provided can only feasibly be provided as a public good, price negotiation *ex ante* is not feasible, the scheme is mutually beneficial for participants and is beneficial for good-recipients, and costs are distributed in proportion to benefits—the principle of fairness generates obligations for those who benefit from the scheme. To avoid fulfilling this obligation would be to free-ride on the scheme, unfairly taking advantage of the cooperative sacrifices of others.

We must now address Simmons's argument that most members of actual political societies cannot plausibly be said to accept the benefits that their cooperative schemes shower upon them. As I will show, Simmons's position comes down to the claim that, because an individual's true preferences, no matter how eccentric, determine what is not valuable to her, the benefits produced by public-goods schemes can fail to be worth their costs so easily that the principle of fairness cannot plausibly serve as a ground for political obligations. If Simmons is right, fair play theories of political obligation are in trouble. However, his argument rests on an implausible account of how we ought to weigh opportunity costs. Once this has been revealed, we will see that there are more attractive alternatives available that will open the possibility of being able to reliably determine, in some cases, when benefits are worth their costs to individual recipients. Let us turn, then, to Simmons's notion of acceptance of benefits.

2. Accepting Benefits Willingly and Knowingly

According to Simmons, a public good must be accepted willingly and knowingly before it can give rise to obligations of fair play. To be obligated to contribute to a public-goods scheme—to be a free-rider for failing to contribute—requires an intention to take advantage of

¹³ Notice that requiring burdens to be proportional to benefits does not require that everyone benefits equally. In my view, such a requirement would be unnecessarily demanding.

those already making sacrifices within the scheme. But this bad intention, Simmons argues, cannot be present unless the received benefit is also accepted. So, the question is, what does it mean to accept a public good? The notion of acceptance is easy to characterize for excludable or optional goods. To accept a benefit of either of these kinds is to “have tried to get (and succeeded in getting)” it (Simmons 1979, 129). But when a good is non-excludable and non-optional, one cannot ‘try’ to get the benefit since one cannot avoid it. It is in these cases that Simmons says a benefit is accepted when it is taken willingly and knowingly.

Taking a benefit knowingly means that one has some understanding of the benefit one receives as proceeding from the productive efforts of others (call this the *knowledge* condition). Taking a benefit willingly means that one does not regard it as having been forced upon one against one’s will, and does not regard the benefit one receives as being outweighed by the cost demanded for it (call these the *no forcing* and *perceived benefit* conditions respectively). According to Simmons, a benefit is accepted when all three of his conditions—the *knowledge* condition, the *no forcing* condition, and the *perceived benefit* condition—are met (Simmons 1979, 132). I will now consider each condition in turn, beginning with the knowledge condition.

According to the view I defend, the principle of fairness generates moral obligations. Given this, it is worth asking whether these obligations are enforceable (by ‘enforcement’ I mean the deliberate imposition of costs on the wrongdoer intended by those who impose them to punish her wrongdoing and to bring her future behavior in line with the violated norm). I will assume for the sake of argument that all moral obligations are enforceable, but that enforcement measures must be proportional to the seriousness of the violated obligation.¹⁴ If I have an obligation to stay off the grass of the campus green, but don’t, I am liable to enforcement

¹⁴ In adopting this view, I follow Arneson (2013).

measures. But if walking on the grass isn't a big deal, enforcement measures should be mild. Those nearby might give me dirty looks, but any punishment more severe than that would be disproportionately harsh.

Often, we will be interested in cases that call for legal enforcement. These are enforcement measures that make use of the apparatus of the government and are often relatively costly. Giving someone nearby a dirty look is cheap and mild, but enforcement processes that involve enacting, administering, and adjudicating laws will often be expensive and labor-intensive. And even if the punishments dispensed through these systems are relatively minor (take small fines, for example), the mere fact of being disciplined by an institution as large and powerful as a government may be a significant cost in itself. Thus, legal enforcement measures should therefore only attach to wrongdoings (or patterns of wrongdoing) that are comparatively serious. Otherwise, the enforcement measures, and their associated costs, will be out of proportion to the wrongs they are meant to redress. In the context of the principle of fairness, my argument will be that proportionality constraints will often prohibit enforcing obligations of fair play, and will almost always prohibit legal enforcement, unless the knowledge condition is satisfied.

Once we adopt the knowledge condition for enforceable obligations of fair play, only the no forcing and perceived benefit conditions will be left as the remaining possible loci of disagreement between Simmons and his opponents. Between these conditions, I will show that only the perceived benefit condition can do any theoretical work for Simmons. I will then give a clarifying account of one of Simmons's arguments in terms of opportunity costs that will allow me to more precisely characterize the kind of relationships one may have to a public-good-producing scheme. To introduce the argumentative significance of the knowledge condition, we

may turn to a debate between Simmons and Arneson over the importance (or not) of accepting benefits.

2.1 Adopting the Knowledge Condition

The fundamental task in discussions of the principle of fairness, recall, is to correctly identify free-riders. Simmons and Arneson agree that bad intentions are part of what it is to be a free-rider,¹⁵ but Arneson objects that Simmons is “looking for a faulty intention or state of mind in the obligated person in the wrong place” (Arneson 2013, 142). Simmons is looking for individuals who genuinely understand the benefits they receive, regard them as worth the price demanded for them, and still refuse to pay. Finding it implausible to think that many real people fit this description, he concludes that very few have obligations of fair play. But Arneson believes we should ask, “Once the true situation is explained to the passive beneficiary...what will she do then” (Arneson 2013, 142). Arneson goes on to argue that, after being informed about the scheme and its benefits, there are only three possible cases in which the passive beneficiary will not contribute to the scheme, and that in all three she has an obligation of fair play to participate.

Before looking at Arneson’s three cases, we should pause to ask why one should bother informing the beneficiary of the scheme from which she benefits before enforcing her duties to contribute to it. It might be easier to just extract compensation, and one might think (at least initially) that there is no weighty moral reason to inform her, even if it would be a nice thing to

¹⁵ As I have argued, my own view is that the bad intention in question (which I have called the free-rider motive) is not a necessary component of being a free-rider. This motive, in the presence of an appropriately produced public good, is sufficient to show that the benefit received is worth the cost demanded, thereby guaranteeing that the non-participant has obligations under the principle of fairness, but it is the value of the benefit and the sacrifices needed to produce it that do the moral work, not intentions.

do. The uninformed beneficiary (who I will refer to as ‘blissfully ignorant’) does, by hypothesis, benefit from the appropriately cooperative efforts of others, and thus owes them compensation in the form of participation in their scheme. If this is the case, and the blissfully ignorant beneficiary is violating her obligations by failing to participate, then why not enforce those obligations straight away?¹⁶

Consider enforcement against the uninformed; must the blissfully ignorant non-participant be informed of the scheme from which she benefits before her obligations are enforceable? The answer, I believe, will almost always be Yes, especially where legal enforcement is concerned. In the case of the blissfully ignorant, to enforce without informing would amount to sending debt collectors to her door, demanding payment without explanation at the threat of violence for unidentified ‘benefits received’. After such an experience she might be more appropriately referred to as the ‘terrorized ignorant’. Of course, we should not forget that sanctions must be proportional to the wrong committed, and, as Arneson observes, sanctions can be very mild (Arneson 2013, 144). But even if the blissfully ignorant person is subjected to no more than dirty looks from her neighbors, she merely goes from a terrorized ignorant to a bullied ignorant. As Simmons insists, “it is the responsibility of those who cooperate to produce open or public goods to inform or otherwise make clear to consumers their expectations of reciprocation” (Simmons 2001, 33).¹⁷ Otherwise, scheme participants will not have put themselves in a position

¹⁶ Arneson, for instance, holds that all violations of obligations are open to proportional sanctioning, though he does not give an account of how to make proportionality calculations (Arneson 2013, 144).

¹⁷ As noted earlier, beneficiaries may sometimes have a duty to inform their benefactors of their benefaction. Since Simmons argues that the principle of fairness can only apply to schemes in which there is genuine cooperation, he effectively rules out cases in which an obligation-generating benefit is accidentally or otherwise unknowingly produced. If beneficiaries have good reason to believe that their benefactors are unaware of the good they produce, beneficiaries will have a duty to inform them. But once this duty is discharged, it will still fall on the participants to the scheme to seek compensation for their productive sacrifices if they so choose.

to rightfully demand payment even if beneficiaries would be willing to pay if they were informed.

If any doubts remain, we can easily list at least three reasons for maintaining that, usually, obligations may not be enforced against the blissfully ignorant unless she is informed of the scheme from which she benefits. First, as I have just argued, enforcement against the blissfully ignorant would be to unjustly terrorize (or at least bully) her. By coercively collecting fees without making a good-faith attempt to explain why, easily preventable psychological harms would be imposed on her. More than this, failing to inform beneficiaries when it would be very easy to do so fails to treat them with the respect and dignity that is due them as our moral equals. As we have already seen in the discussion of price negotiation, everyone is owed this kind of respect.

Second, if the scheme is encoded in the law, the rule-of-law norm of notice requires that obligors be informed of the law before it is enforced against them. We would not want to follow in the footsteps of Nero, who posted laws at the top of tall poles where no one could read them while still punishing those who failed to comply (Scalia 1997, 17). And third, allowing such practices would open the door to terrible abuses in the name of providing for the ‘general welfare’. It is, after all, not just beneficiaries who might want to unfairly take advantage of others. If those who receive benefits need to be held accountable, then so do those who purport to produce them. In a proportionality calculation, all of these considerations will weigh heavily against enforcement against the blissfully ignorant.

If my argument above is correct, then the best way to read Arneson when he says that we must ask what a passive recipient *will* do once she *is* informed of the benefit-providing scheme (as opposed to what she *would* do if she *were* informed), is that cooperators must make a good-

faith effort to inform her of the scheme before collecting dues.¹⁸ So it is not mere receipt of benefits that generates enforceable obligations under the principle of fairness, but *informed* receipt of benefits. A straightforward way to reconcile this conclusion with the intuition that all breaches of obligation warrant some kind of sanction is to say that the ‘sanction’ due the blissfully ignorant free-rider is precisely that she be informed of the scheme from which she is benefiting. Receiving and understanding the explanation will not be free, costing her at least some time and energy as well as her future participation. Granted, this sanction is very mild, but once the ignorant beneficiary is informed she will be fully obligated to contribute her fair share within the scheme from which she benefits.

2.2 Is Non-Culpable Ignorance Blissful Ignorance?

We may now return to the argument of Arneson’s I brought up at the beginning of section 2.1. Arneson argues that once a passive recipient of benefits is informed of the scheme from which she benefits, there are only three reasons for which she might refuse to participate. First, she may fully understand the scheme and its benefits, believe that it is worth its cost to her, but still refuse to contribute. In this case, she is obviously acting on a free-rider motive. Second, she may remain ignorant of the workings of the scheme or the value of its benefits to her, but culpably so. Since culpable ignorance clearly cannot excuse her from her obligations, she counts as a free-rider in this case as well. Third, she may remain ignorant (as in the second case), but non-culpably so. For whatever reason, she non-culpably fails to understand the scheme or the value of the benefits it provides. Note that in this case the benefactor is not blissfully ignorant. In

¹⁸ Arneson claims (contra Simmons) that acceptance of benefits need not be willing and knowing. Instead, he says, “unwilling acceptance of benefits will do” (Arneson 2013, 142). This phrasing at least leaves room for the claim that acceptance must be knowing before obligations arise, even if it is not willing.

a state of blissful ignorance, the beneficiary really knows nothing about the scheme, usually because benefactors have not informed her of it. The question is, even in the case of non-culpable ignorance (as opposed to blissful ignorance), is the beneficiary obligated to contribute to the scheme?

Arneson believes the answer is Yes. The non-cooperator's "actual relations...to the cooperative behavior of her fellow citizens generate a reciprocal obligation under the...principle of fairness" (Arneson 2013, 142). However, we might worry that the informed but non-culpably ignorant person is not relevantly different from the blissfully ignorant person. I will argue, however, that the two cases are distinct.

Arneson notes that the non-culpably ignorant person is mistaken either about the structure of the scheme or about the benefits it yields. The blissfully ignorant person, however, does not have enough information even to be mistaken about the scheme; she has made the more basic mistake of failing to believe there is a scheme in the first place. In the case of the informed but non-culpably ignorant person, she is likely to understand that there is supposedly a scheme that produces benefits that are valuable to her, and she is likely to have at least some grip on how the scheme is supposed to work and what kinds of sacrifices others are already making. The problem is that she undervalues the benefits, mistakenly believes that the scheme is unjust, etc. Thus, while it will be onerous for her to be coerced into contributing, she will not be wronged by the coercion as the blissfully ignorant person would be.¹⁹ Even so, in rare cases, attempts to inform a

¹⁹ A question should arise here. I have said before that the compensation owed under the principle of fairness is either what one would have agreed to had *ex ante* price negotiations been possible, or a fair share of the costs of the scheme, whichever is lower. In the case just described, it seems clear that the beneficiary would not have agreed to pay anything, so why think she must contribute? I am relying here on the claim that in certain cases an individual may be mistaken about how valuable a good is to her. I will have more to say in defense of this claim when I get to Simmons's perceived benefit condition below.

We must also not forget that contributions are owed in proportion to the value of the benefits received. The onerousness of being forced to contribute to a scheme might legitimately make the benefit received less valuable, and thus reduce the compensation owed. Interestingly, this parallels a debate in the literature on compensation owed

blissfully ignorant individual might absolutely fail. If the good-faith efforts on the part of the blissfully ignorant person to understand the good-faith efforts of scheme-participants to inform her genuinely fail, and because of this failure she cannot leave her state of blissful ignorance, she almost certainly acquires no enforceable obligations of fair play.

One might object that, sometimes, duties of compensation do arise even when communication is impossible (i.e. when the wrongdoer cannot leave a state of blissful ignorance). The argument I have given above relies primarily on the claim that the harm done to the blissfully ignorant beneficiary in enforcing her obligations would be disproportionately large relative to whatever benefits would accrue to other contributors if she participated. But this need not be the case in principle. Consider an analogous case in which an elderly gentleman sees a chair on roller wheels sitting by the curb near a trash can and justifiably concludes that its owner has left it there to either be taken by a passerby or by the garbage collectors. He takes the chair home, breaks it down for parts, and then completes his intricate lawn sculpture with it. Immediately thereafter he suffers a stroke that prevents him from understanding language, written or spoken.

The chair's owner, however, had not discarded it. She had left it by her moving truck and an unusually strong gust of wind had rolled it down the hill and left it by the trash cans. When the chair owner hears what happened from a neighbor, she goes to the elderly man's home to claim compensation (perhaps money) equal to the value of the now dismantled chair. Unfortunately, the elderly man cannot be informed of the situation because of his stroke and will be distressed when compensation is seized, thinking he is being robbed. So here we have a case

by third parties who unavoidably and innocently benefit from harms to others (cf. Fullinwider 1975; Butt 2007; Butt 2014).

in which an individual innocently and unknowingly benefits at some cost to another in a way that would, under normal circumstances, give rise to a duty to compensate.

May the chair owner claim compensation for her chair? The answer seems to depend on how stringent the elderly man's duty to compensate is, where this stringency seems to me to depend on two main variables. The first is the harm that would be imposed on the victim (the chair owner) if compensation were not rendered, and the second is the harm that would be imposed on the victimizer (the elderly man) if compensation were rendered.²⁰ Clearly, depending on how we flesh out the example, seizing compensation for the chair may or may not be permissible. Suppose the elderly man took a cheap and worn-out desk chair and would be extremely distressed at compensation being taken. Here I think the chair owner may not recoup her losses. Alternatively, suppose the chair was a rare antique and selling it was the owner's only opportunity to get enough money for an expensive medicine to save her child's life. Furthermore, the elderly man is calm and collected and would not be very distressed at compensation being seized. In this case I think the chair owner may claim compensation. In many other variations of this example, when the weightiness of the two variables is close to equal, the permissibility of reclaiming the chair will be unclear. In all these cases, the elderly man represents those members of a political society unable to leave a state of blissful ignorance with respect to the public-goods schemes from which they benefit. And we want to know whether they, like the elderly man, may sometimes be coerced into participating in those schemes despite their blissful ignorance.

Returning to the principle of fairness, the harm caused by an individual who fails to submit to the rules of a cooperative scheme instituted across an entire large-scale political society

²⁰ To clarify, the harm to the victimizer in question does not include the loss of the value of the chair since he had no right to it to begin with. Rather, the harm in question is the distress and felt disrespect involved in seizing the chair without explanation, causing the elderly man to reasonably believe he is being robbed. Also, though I do not discuss the possibility here, partial compensation should be considered when it is viable and appropriate.

will usually be very small. Thus, when communication is impossible, duties of fair play to compensate participants of public-goods schemes will rarely be stringent enough to outweigh the costs of enforcement. We cannot entirely rule out the possibility of cases in which enforcement against the blissfully ignorant is permissible, but the harms imposed on the blissfully ignorant by enforcing their duties of fair play will usually outweigh the harms imposed on other participants by their failure to participate.

What I have argued is not that benefits must be received knowingly before obligations of fair play arise. Rather, benefits must (almost always) be received knowingly before obligations are properly enforceable, at least in the context legal enforcement within a large-scale political society. While this is not strictly consistent with Simmons's acceptance account of fair play, I think it coheres with the spirit of his argument, at least with respect to the knowledge condition. What remains, then, is to consider the no forcing condition and the perceived benefit condition, the two of which constitute Simmons's notion of taking benefits willingly.

2.3 Simmons's Preference-Based Theory of Value

Simmons's willingness condition on acceptance of public goods requires that each recipient not regard the benefits she receives as having been forced upon her against her will (no forcing), and that she must regard the benefit as being worth the price she is required to pay for it (perceived benefit). It might at first appear as though these conditions do not get us very far in determining where Simmons and opponents like me or Arneson part ways. Everyone agrees that the benefits provided by a scheme must be worth their cost to recipients before they become obligated to contribute, and the no forcing condition looks like nothing more than a paraphrase of

the voluntarist claim that fair play obligations cannot be acquired non-voluntarily.²¹ So at first glance, neither seems able to do much argumentative work for Simmons. But the apparent agreement about the perceived benefit condition is misleading. The important question to ask here is, what makes a benefit valuable?

On the face of it, Simmons's answer to the question above is relatively straight-forward. When he is explicating what he means by 'willingly and knowingly,' he says that acceptance "involves a number of restrictions on our attitudes toward and beliefs about the open [public] benefits we receive" (Simmons 1979, 132). So, for his three acceptance conditions to be satisfied, it is necessary and sufficient to have certain kinds of beliefs and attitudes about the benefits accepted. I should note, however, that it would be possible for Simmons to maintain that having these beliefs and attitudes is merely necessary for acceptance, but not sufficient. For instance, he could argue that for a benefit to be valuable to a recipient, she must both believe that it is valuable to her, *and* she must be right according to some additional set of criteria. Simmons does not seem to consider a view like this, and his examples always focus on the preferences of beneficiaries. Furthermore, he argues that theorists like Klosko who often seem centrally concerned with the importance of particular public goods such as security are unlikely to be presenting fairness-based theories at all (Simmons 2001, 35-36).

Whatever Simmons's larger theory of wellbeing is, he at least maintains that a benefit is *not* valuable to a recipient unless she *regards* it as valuable. Thus, Simmons can argue that even relatively efficient and just schemes that produce 'obviously' valuable goods can fail to generate

²¹ And in fact I think that is all it is. Simmons never appeals to the no forcing condition as part of a substantive argument in either his 1979 chapter or his 2001 chapter, and clearly he cannot since a central point in question is whether or not special obligations can be acquired non-voluntarily. In his 2001 chapter (p.36) Simmons constructs a case in which a benefit is forced upon a recipient against his will, but Simmons only uses this fact to more emphatically conclude that the benefit could not possibly have been worth its cost—i.e. that the perceived benefit condition is all the more certainly violated.

obligations under the principle of fairness. All the recipient must do is honestly regard the benefit received as not worth its cost. So we can, I think, safely say that on Simmons's view, a person's true preference orderings determine what is *not* valuable to her. Simmons could allow that certain things can be dis-valuable to a person regardless of her beliefs and attitudes, but what is important for the argument here is that if a recipient regards what she receives as dis-valuable to her, then it is.

What should be said about the possibility of agents having imperfect access to their own preference orderings? Simmons can easily concede this possibility while maintaining that what makes the benefit valuable for the individual recipient is her own preference orderings. We might say that, as a theoretical matter, getting at the relevant preferences requires taking the beneficiary, giving her perfect access to her own preferences, and making her perfectly instrumentally rational. This process would still permit her to have any preferences at all, and Simmons acknowledges this. He writes, "An individual's preferences may...be unusual or eccentric; but provided these preferences are not based in negligent belief or ignorance, this hardly makes it any less wrong to impose on her the burdens associated with others' schemes, of which she wants no part" (Simmons 2001, 35). A person may genuinely prefer danger, sickness, hunger, and thirst to safety, health, food, and drink, and so long as those truly are her preferences it is not unfair for her to refuse to contribute to a scheme that publicly provides these goods since she does not benefit from them. So, for Simmons, the value of a benefit to a recipient is always dependent on her preference orderings.

2.4 Opportunity Costs: Free-Riders and Efficiency

On a view of the kind Simmons seems to hold, any preference ordering might be genuine, so there are no goods, public or private, that can be assumed to be valuable to each and every person. But Simmons, I think, wants to say more than this. In order to convincingly defeat the claim that the principle of fairness obligates most people to obey the law, Simmons argues that it is not only possible, but easy for public-goods schemes to fail to be worth their cost, even in the case of schemes providing essential goods such as fresh water. To show this, he asks us to consider a rural village in which wells are going dry. One resident, call him Will, wants to dig a new, deeper well that will provide fresh water for himself and his family. The other villagers, however, want to divert a nearby river so that it flows through the village and provides fresh water to all. Will makes all the necessary preparations to dig his well, but the day before he can begin digging the other villagers run their trench through the only place where he could have dug it.²² Simmons writes:

While I [i.e. Will] have benefitted in significant ways from the cooperative sacrifices of others, there is something deeply unconvincing about the claim that I *owe* my neighbors reciprocation for essential goods that they, in effect, *forced* me to take from them, denying me the option to provide the goods for myself or to do without. Surely, in such a case, only if I *prefer* benefitting from the cooperative scheme to benefitting from self-provision (or to doing without) could I be accused of unfairly taking advantage of my neighbors when I refuse to do my part in the scheme. (Simmons 2001, 36, original emphasis)

While some of Will's preferences are satisfied—he prefers fresh water to no water—they are not satisfied in the *way* he most prefers. So, Will's highest ranked bundle of preferences (fresh water via the well) is not satisfied, and he is forced to live with the satisfaction of a lower ranked

²² Simmons does not specify what options are open to the trench-diggers, but let us suppose that for whatever reason the trench could not feasibly have been dug along a different course. Otherwise, intuitions about the case become unclear since, presumably, the trench-diggers wrong Will if they take away his ability to dig a well unnecessarily. I should also note that the trench does not provide a public good strictly speaking since taking water from it is optional, but this can be overlooked for our purposes here.

bundle of preferences (fresh water via the trench). Because of this privation, Simmons concludes that if Will refuses to do his part in the trench scheme he will not take advantage of his fellow villagers. If Simmons's analysis is right, public-goods schemes fail to be worth their costs very easily since recipients of benefits need only prefer some alternate means of provision. But Simmons's conclusion is too quick.

The core of Simmons's claim is that the opportunity costs of a scheme must be weighed against its benefits. Taking up this idea, I will first argue that once opportunity costs are explicitly incorporated into the analysis, we will be able to distinguish three kinds of recipients of goods: full beneficiaries, reluctant beneficiaries, and reluctant recipients. As we will see, full beneficiaries and reluctant recipients constitute the two extremes of benefiting maximally from a scheme, and benefiting not at all. Reluctant beneficiaries, however, receive variable benefits. Depending on the opportunity costs they bear, they will owe more or less to a scheme from which they benefit. It is this middle category of reluctant beneficiaries that Simmons's argument does not take into account, and that ultimately robs it of its strength. Finally, I explain how an understanding of opportunity costs allows us to more precisely account for the badness of inefficiency.

Simmons's move is to factor into our cost-benefit calculations not just the costs of participating in the established scheme, but the opportunity costs of foregoing alternative forms of provision. The idea seems to be that if the value to Will of provision via well is greater than the value of provision via trench, then adding the lost value of the well to the costs of the trench scheme guarantees that the trench scheme is not worth its costs. However, we can separate the good provided from the means of provision. Suppose, for instance, that having plenty of fresh water easily available is worth 50 utiles to everyone in the village however it is provided. For

Will, provision via well is worth another 50 utiles, while provision via trench is worth 25. For the other villagers, provision via trench and well are worth 50 and 25 utiles respectively. Thus, the trench scheme is worth 75 utiles to Will while the well would have been worth 100 utiles to him. Factoring in the loss of 25 utiles as a cost of the trench scheme to Will, Will only gets 50 utiles from the trench scheme instead of 75, while the other villagers get 100 utiles. Suppose now that the costs of maintaining the trench scheme are fairly low, and they are distributed so that Will is asked to contribute 1 utile to its maintenance while everyone else contributes 2 utiles. Since Will benefits half as much as others, and contributes half as much as others, burdens are distributed in proportion to benefits. Also, the benefits Will receives far outweigh the burdens he must bear. Thus, there seems to be no problem here with saying that Will benefits sufficiently to make failure to contribute free-riding.

According to the picture sketched above, it is the *marginal*, not total, value of the foregone opportunity that ought to be subtracted from the value of a scheme. It makes no sense to subtract the value of readily available fresh water from the trench scheme when that is exactly the good it provides. Of course, if the two schemes provided *different* goods, the story would be different. If the other villagers diverted a river of chocolate fondue instead of a river of water, it would be perfectly legitimate to subtract the full value of the foregone fresh-water-providing well from the value of the chocolate-fondue-providing trench. The most immediate conclusion, then, is that it is not as easy as Simmons supposes for the provision of a public good to fail to be worth its costs. But more than this, we are now in a position to make much more fine-grained distinctions among those who wish not to participate in a cooperative scheme.

So far, I have referred to those who unfairly fail to contribute to a cooperative scheme as free-riders. Now, however, we can give a more detailed account of what is owed to a scheme for

failure to contribute. First, we have what we might call full beneficiaries. These individuals want the good that is provided, and they want it in the way that it is provided. Thus, they are not negatively affected by opportunity costs and owe a full fair share to the scheme from which they benefit. If a full beneficiary fails to contribute to a scheme, she is a classic free-rider. Assuming the knowledge condition is satisfied, she must be moved by the free-rider motive and owes not only her fair share of the cooperative burdens, but also reparations for her malicious attempt to take advantage of the cooperative efforts of others.

Will, however, is not a full beneficiary. He falls into an intermediate group that we might call reluctant beneficiaries. These individuals do benefit from the cooperative scheme in question, but because the good is provided suboptimally (relative to their preferences) they benefit less than they otherwise might. These individuals do owe their fair share to the scheme from which they benefit, but the size of this share varies inversely with the opportunity costs imposed by foregoing other available means of provision.

Finally, we have the third class of individuals, who we can call reluctant recipients. These individuals would rather forego the good entirely than receive it at any level of cost. We might say that, for these people, the opportunity cost of giving up e.g. free time swamps whatever benefits the scheme provides. Since these individuals do not receive a net benefit from the provision scheme, they are not obligated to contribute to it.

My last point with respect to opportunity costs is that the analysis I have given allows us to see more clearly what to say about inter-scheme inefficiency. At the end of section 1.1 I said that a cooperative scheme can be significantly inferior to an available alternative if that alternative is significantly more efficient.²³ Then, if the inferior scheme is instituted, it does not

²³ Here I do not attempt to address the question of how to go about identifying available alternatives. All my argument needs is that available alternatives exist.

generate obligations under the principle of fairness. This is because the opportunity cost of foregoing the efficient alternative reduces the compensation each beneficiary owes to the scheme. And when the opportunity cost is large, no one (or almost no one) receives a net benefit from the scheme and thus owes no compensation. Alternatively, if beneficiaries do owe some compensation, though only a little, the scheme may not be able to cover its costs, or enforcement might violate proportionality constraints. But if a scheme is approximately as efficient as available alternatives, opportunity costs at the margins will not prevent the scheme from being efficient enough.²⁴

3. Conclusion

Significant strides have been made toward defending the principle of fairness against Simmons's acceptance-based arguments. I have argued that of the three conditions that constitute his notion of willing and knowing acceptance of benefits, only the perceived benefit condition can do the work of separating his view from those of other theorists who maintain that benefits need not be voluntarily accepted for obligations of fair play to arise. Thus, setting aside empirical and theoretical questions about whether or not modern political communities engage in properly cooperative enterprises, the controversy between voluntarist and non-voluntarist formulations of the principle of fairness comes down to this one disagreement.

Once we recognize opportunity costs as a legitimate kind of cost a scheme may impose on beneficiaries, new possibilities emerge. There are not simply beneficiaries and non-

²⁴ Another kind of inefficiency arises when a scheme is as good as available alternatives, but still produces some goods that can't be fairly distributed. Arneson gives the example of a community with a cistern that it uses to store its water. There is a distribution scheme to prevent overuse, but the top little bit of water will evaporate if unused (Arneson 2013, 140). The question is whether it is permissible for a community member to use that little bit of water over and above her usual allotment of water since it will evaporate anyway. My own sense is that it is not permissible to take the water unless one contributes a little bit more to the scheme to maintain a distribution of costs in proportion with benefits.

beneficiaries, but benefits and burdens come in degrees, varying not just with the value of the benefits themselves, but also with value of the benefits that the scheme crowds out. While this result allows us to reject Simmons's claim that schemes can very easily fail to be worth their costs, it makes worries like Simmons's "baseline problem" all the more pressing (Simmons 2001, 37-38). The problem, according to Simmons, is that of determining which baseline conditions we should appeal to when we determine whether or not a scheme is beneficial. Similarly, we must now worry about what determines the "available alternatives" that are foregone when a particular scheme or bundle of schemes is instituted or maintained. Even so, as long as the benefits provided are large, we should be able to plausibly maintain that the opportunity costs of alternative forms of provision account for, at most, relatively minor opportunity costs.

The principle of fairness, as I have defended it, should now be more clearly apt for application to political societies. At least for important goods, we can maintain with relative confidence that their provision as public goods generates obligations of fair play in recipients to compensate those who have already submitted to the rules of the provision scheme. Furthermore, extension of the principle to arenas like international law or environmental protection can now be examined with greater precision. The basic question is no longer whether or not e.g. a given state is a full beneficiary of international law, but whether or not the current system of international law is outweighed by nearby alternatives. Thus, in light of the claims defended in this chapter, the principle of fairness should be seen as a valuable resource for accounting for political obligations, obligations to obey international law, environmental obligations, and more.

Chapter 1, in full or in part, is currently being prepared for publication of the material. Finley, Aaron. The dissertation author was the primary author of this material.

References

- Arneson, Richard. 2013. "Paternalism and the Principle of Fairness." In C. Coons and M. Weber (Eds.), *Paternalism: Theory and Practice*, 134-56. Cambridge: Cambridge University Press.
- Butt, Daniel. 2014. "'A Doctrine Quite New and Altogether Untenable': Defending the Beneficiary Pays Principle." *Journal of Applied Philosophy* 31 (4): 336-48.
- Butt, Daniel. 2007. "On Benefiting from Injustice." *Canadian Journal of Philosophy* 37 (1): 129-52.
- Dworkin, Ronald. 2011. *Justice for Hedgehogs*. Harvard University Press.
- Fullinwider, Robert. 1975. "Preferential Hiring and Compensation." *Social Theory and Practice* 3 (3): 307-320.
- Goodin, Robert. 1989. "Theories of Compensation." *Oxford Journal of Legal Studies* 9 (1): 56-75.
- Hart, H. L. A. 1955. "Are There Any Natural Rights?" *The Philosophical Review* 64 (2): 175-91.
- Klosko, George. 2004. *The Principle of Fairness and Political Obligation*. New Edition. New York: Rowman & Littlefield Publishers.
- Klosko, George. 2019. *Why Should We Obey the Law?*. Medford: Polity Press.
- Ostrom, Elinor. 2010. "Beyond Markets and States: Polycentric Governance of Complex Economic Systems." *American Economic Review* 100: 641-72.
- Rudd, Anthony. 2007. "In Defence of Narrative." *European Journal of Philosophy* 17 (1): 60-75.
- Scalia, Antonin. 1997. *A Matter of Interpretation: Federal Courts and the Law*. Princeton, New Jersey: Princeton University Press.
- Simmons, A. John. 2001. "Fair Play and Political Obligation: Twenty Years Later." In *Justification and Legitimacy: Essays on Rights and Obligations*, by A. John Simmons, 27-42. New York, New York: Cambridge University Press.
- Simmons, A. John. 1979. "The Principle of Fair Play." In *Moral Principles and Political Obligations*, by A. John Simmons, 101-142. Princeton, New Jersey: Princeton University Press.

Chapter 2

Fair Play and Artificial Agents

Introduction

As the previous chapter revealed, a core task in applying the principle of fairness (PoF) is evaluating the value to a person of the benefits she receives. At the individual level, the debate is about (a) what makes something valuable to an individual, and (b) when we may permissibly restrict an individual's liberty. The PoF says that we may restrict the liberty of free-riders, and then characterizes free-riders as those who receive benefits that are worth their cost *according to the right metric* without contributing their fair share. So, the fundamental question is, what is the right metric? According to A. John Simmons, the metric for each human is the one set by her preferences. When we transition from humans—natural agents—to artificial agents such as states or corporations, must we adopt a new metric? In this chapter, I argue that the answer is yes.

My starting point is a paper by David Lefkowitz in which he attempts to extend the PoF to international law. He argues that we can take the principle as Simmons defends it and apply it, unaltered, to artificial agents. The core of his argument is the claim that officials of the state, acting in their official capacity, can be treated as the state itself. Thus, when we apply the principle to states, we really apply it to individual humans who are acting on behalf of the state. Importantly, this means that a state voluntarily accepts the benefits of a public-goods scheme when the relevant official(s) see those benefits as worth their cost to the state. However, we have good reason to be skeptical of this strategy.

As we will see, Simmons's version of the principle makes demands on the mental states of beneficiaries that Lefkowitz's system cannot adequately accommodate. The most pressing problems for cost-benefit calculations arise when we are required to aggregate the mental states

of a body of officials, a legislature for instance, who may act as they do for a wide variety of reasons. Passing an international treaty into law, for instance, does not imply that most legislators see it as worth its cost to the state. I will argue that appealing to the state's preferences is the wrong way to approach these calculations. Rather, I will argue, we should measure the value of benefits received by a state against a perfectionist notion of state well-being.

The chapter proceeds as follows. In section one, I will briefly motivate the idea that the PoF should be extended to artificial agents and set the focus of the discussion on moral obligations rather than international law. In section two, I show that the cognitive constraints of the PoF make its application to artificial agents complex in ways Lefkowitz notices but fails to address. Thus, we need a different way to measure costs and benefits to artificial agents—we need an account of their well-being. In section three, I will argue that we should adopt a perfectionist account of state well-being which allows us to make cost-benefit calculations for the state without appealing to thick mental states of the state. In section four, I argue that state well-being is at most instrumentally valuable and introduce the notion of *alignment* to describe the circumstances in which its well-being takes on moral significance. This approach to state well-being allows us to answer standard objections facing perfectionism as an account of individual well-being. Finally, in section five, I argue that state obligations under the PoF can be causally important in generating obligations for individuals even though state well-being has at most instrumental moral value.

1. Mediterranean Migrants and International Law

If collective artificial agents do have well-being, and can bear obligations, questions arise about the relationship between artificial agents and the natural persons that inhabit or constitute

them. Does the well-being of the artificial agent have any intrinsic moral importance, or is its significance purely instrumental? Similarly, does it have obligations that are not mere aggregates of the obligations of its constituents? If so, do these obligations devolve to constituents? I will answer each of these questions in turn over the course of this chapter. In this section, however, I begin with an example to draw out the intuition that states can indeed bear obligations toward one another and focus the discussion on moral questions rather than on questions of international law.

In 2013, a makeshift vessel carrying upwards of 500 migrants and refugees caught fire and sank off the coast of Lampedusa, a small Italian island in the Mediterranean south of Sicily. More than 300 of those aboard died. The tragedy was widely publicized and in response Italy began Operation Mare Nostrum in late 2013 in which its navy patrolled portions of the Mediterranean to rescue distressed vessels. Despite the success of the operation (it is credited with rescuing more than 200,000 migrants during the year it operated), Italy felt overburdened by its costliness and called on the E.U. for support. In response, the E.U. border agency Frontex began Operation Triton to replace Operation Mare Nostrum. Unlike Mare Nostrum, Triton operates relatively close to European shores and prioritizes border control over search and rescue. Thus, Operation Triton has been widely criticized for failing to meet a pressing humanitarian crisis.

In this case, it seems intuitively clear that Italy and other EU nations had a natural duty to rescue distressed migrants. Once Italy started its scheme, other EU nations acquired an additional obligation to help Italy in its efforts. Whether these duties and obligations originate at the collective level, or are aggregated from the individual level, and what the relationship is between collective and individual responsibilities, are matters I will comment on later. For now, to

develop a clean version of the case focused on fairness, I will make several simplifying assumptions. I will assume, contrary to fact, that we have positive duties only to co-nationals.²⁵ On this assumption, Italy's rescue effort was an act of pure supererogation. I also stipulate, for the sake of argument, that Italy's efforts significantly improved Europe's control of its borders, and that this additional control was a significant benefit to each member state.

In the polished EU case, Italy's actions establish a search and rescue scheme that, by hypothesis, produces benefits that are worth their cost for other member states.²⁶ This makes the polished case apt for fairness obligations. More generally, when an artificial agent that acts to produce benefits that fall on other artificial agents, we want to know whether these agents as such come to have fairness obligations.

There is no harm in admitting that artificial agents can have obligations qua artificial agents. The essential capacity they must have is the ability to recognize and act on moral reasons. If they can do this, they can have duties and obligations. Additionally, if intelligible calculations can be made about their well-being, they can have fairness obligations. There is no harm in allowing all this because it does not immediately follow that these duties and obligations (and maybe rights) have any moral weight. Just because a state has a duty doesn't mean that duty matters morally, nor does it mean that the individuals who constitute the state have an obligation

²⁵ I drop this assumption in chapter four where I develop the relationship between natural duties and fairness obligations. Both here and there I take it for granted that we do have positive duties of beneficence. Given this, it is implausibly arbitrary to think that they extend only to co-nationals.

²⁶ Given my assumptions, the transition from Operation Mare Nostrum to Operation Triton is sensible. If the only relevant moral imperative is to efficiently share the costs and benefits of border control, and if Triton protects the border as well as Mare Nostrum, but at a lower cost, then the switch replaces one scheme with a more efficient one. In this case, as an additional consequence, many fewer migrants are rescued. This is bad, but not impermissible according to the assumptions. What this demonstrates is the familiar point that promoting one's own well-being is not always morally permissible.

The fact that the migrants are apparently left out of the scheme may seem odd. After all, they certainly benefit. But not all benefits generate fair-play obligations. The migrants benefit under a natural duty to rescue, not a fair scheme for mutual benefit. For a more detailed discussion, see chapter 5 section 3.3.

to act so as to cause the state to fulfill its duty. The same goes for obligations. In the EU case, for instance, duties to rescue distressed migrants are clearly morally important, while obligations to other states to bear a fair share of the burdens of border-control may not be.

It is worth noting in passing that the EU responded to Italy's fairness appeal with legislation. We can plausibly interpret this response as indicating that the EU recognized a preexisting obligation which it then mobilized through law. Clearly, in the international arena as in the domestic, the principle of fairness can generate obligations independently of law. When it comes to international law, views differ about its sources, legitimacy, and essence, and it is an open question whether or not the principle of fairness could legitimate the international law already on the books. Regardless, what is clear is that, wherever the PoF creates obligations, it can create them independently of the law.

The entire analysis of this chapter rests on the idea that collectives such as states can be obligation-bearing agents. We talk about states and corporations as though they are agents, but we might doubt that this is more than a convenient shorthand for describing the coordinated activities of large numbers of natural agents. So, what does it take to be an agent? I will follow List and Pettit who argue that, at minimum, agency has three criteria. First, an agent must have "representational states that depict how things are in the environment"; second, it must have "motivational states that specify how it requires things to be in the environment"; third, it must have "the capacity to process its representational and motivational states, leading it to intervene suitably in the environment whenever that environment fails to match a motivating specification" (List & Pettit 2011, 20).

These conditions provide a very thin conception of agency—one that even simple robots can satisfy—and states satisfy them easily. States have a great deal of information available to

them and should therefore have no difficulty representing their environment. States also have goals recorded in legislation, official statements, and so on. Thus, they satisfy the motivational condition. Finally, states have the capacity to process and, through their officials, act on their representational and motivational states. Whether or not this capacity is regularly or consistently exercised is beside the point. Even if the capacity for agency were never exercised, so long as the capacity exists, agency exists. Also, the information available to states includes information about what is right and wrong, good and bad. This information can be taken into account in motivational states and can play a role in the way states act. Thus, on this understanding, states are capable of moral agency.

I turn now to Lefkowitz's arguments and his attempts to resolve concerns about the ability of states to have the kinds of mental states demanded by the PoF.

2. Lefkowitz on the Mental States of States

The claim that an entity like a state can be an agent can seem strange. We are used to humans and animals with a kind of independence collective artificial agents do not have. A robot piloted by several humans seems like an odd candidate for agency, much less a robot piloted by one human. The agency of humans and animals is not explainable in terms of other agents, so it is natural to think that any duties or obligations of the robot are really borne by its pilots. But this line of thinking amounts to a genetic fallacy. Human agency is explainable in terms of neurochemistry, social conditioning, and other factors that do not involve the human's own agency, and that sometimes involve the agency of others. If ordinary humans can bear duties and obligations, there is no reason in principle why a robot, or a state, cannot. Aside from this

apparent oddness, additional complications arise in relation to the PoF and its cognitive demands.

Simmons argues that the PoF does not bind unless one voluntarily accepts the benefits produced by a scheme. Lefkowitz claims that Simmons is attempting to reconcile the PoF with the importance of individual liberty. Everyone agrees that free-riding is wrong, and that no one should be subject to the arbitrary will of another. By arguing that free-riding is possible only when I voluntarily accept the benefits of a scheme, Simmons ensures that others' ability to place obligations on me through the PoF is conditional on a voluntary act of my own will. And in this case, the act is the declaration of a subjective preference ordering.²⁷ Finally, these preferences orderings are the metric Simmons argues we should use to determine whether a benefit is not worth its cost to a recipient (Simmons 1979, 128-136). Whether or not one is a voluntarist, the PoF requires some account of well-being with which to make cost-benefit calculations for individual beneficiaries. Lefkowitz, however, follows Simmons in defending voluntarism as the only way to adequately respect the value of liberty, even when we consider artificial agents such as states.

Lefkowitz points to political self-determination as the kind of liberty at stake in extending the principle of fairness to states. Unless states are left free to accept or reject benefits as they please, they will be vulnerable to the arbitrary will of others. The worry is that one state (or several states working together) could drop benefits on another state and then demand payment (cf. Nozick 1974, 93-95). If it is predatory in Nozick's example for my neighbor to thrust a book

²⁷ I say 'declaration' to emphasize Simmons's voluntarism. He often writes as though my preferences are whatever I declare them to be, though we might wonder to what extent this is the case. Preferences are not always subject to agential control or revision, can be mutually incompatible, and don't always represent their possessor's best interests, perhaps even from her own perspective. I do not take up these complications here, but I will argue that they can be smoothed over in the case of collective artificial agents.

into my hands and demand payment, then, the thought is, it should seem just as predatory for several countries to thrust a scheme onto another country and demand compliance. Thus, when Italy starts operation Mare Nostrum, other EU member states are morally at liberty to reject the scheme as not worth its cost. They cannot be bound independently of any voluntary act of willing. Otherwise, says Lefkowitz, Italy can use the PoF to violate the legitimate self-determination of other EU member states by forcing them to join its scheme against their will (Lefkowitz 2011, 335).

As I argued in the previous chapter, however, the debate between voluntarists and non-voluntarists comes down to the question of what makes a benefit valuable to a recipient; if we answer this question differently than Simmons, we can arrive at non-voluntaristic interpretations of the PoF.²⁸ Importantly, we might defend different accounts of well-being for different kinds of agents. Thus, there is no contradiction in principle between being a voluntarist about fairness obligations for natural agents while being a non-voluntarist about fairness obligations for artificial agents. While I think we should be non-voluntarists in both cases, I leave to one side questions about what determines the well-being of natural agents. In the case of states as artificial agents, however, I argue that their preferences play at most a minor role in determining their well-being. Instead, the well-being of the state is dependent on the well-being of its constituents. Thus, for states, the PoF takes on a non-voluntaristic guise. Within this framework, there is room for political self-determination to contribute to a state's well-being, but only as an instrument for deriving value for natural agents. If no natural agents (or too few) benefit from a particular kind of political self-determination, then it is worthless.

²⁸ The labels 'voluntarist' and 'non-voluntarist' are misleadingly binary. One could defend a mixed view in which an individual's voluntary acts of willing do some work in determining the value to her of the benefits she receives, but not all the work.

What, then, of states' mental states? In the case of humans, we appeal to preferences, desires, intentions, and the like in making judgments about the applicability of the PoF. Does the beneficiary want the benefit in the first place? Would she rather have something else? Is she avoiding contributing to the scheme only because she wants to get the good without paying for it? And as Lefkowitz observes, we are quite naturally inclined to speak in similar ways about states. Does Germany want stronger border control? Would Spain prefer a different scheme? Is the UK refusing to contribute only because it wants to dodge the costs of the benefits it receives? The account of agency I have adopted supports the idea that even a simple robot can have thin mental states, but they might be too thin for the PoF. In particular, the notions of benefitting and wanting and preferring might require some kind of subjective experience, experiences which I take it for granted that states and robots do not have.

Lefkowitz addresses this problem by arguing that when we speak of a state's intention to sign a treaty, for instance, we really refer to the intentions of officials within the state, acting in their official capacity, and empowered to sign the treaty in question. This neatly answers any doubts about the mental capabilities of states by having human minds constitute the state's mind. We can call this the *state officials* account of state agency. Lefkowitz's aim is to show that on the state officials account, states can pass Simmons's conditions for voluntary acceptance of benefits just as well as any human. This is plausibly true for the knowledge condition since the relevant knowledge is either in the heads of the officials acting on behalf of the state, or it is at their fingertips.

The state officials account is less plausible for the requirement that beneficiaries not regard the costs of a scheme as outweighing its benefits. According to Lefkowitz, the mental states of the state are supposed to derive from the mental states of the officials within it. Thus, by

definition, there cannot be mismatches between the two. So, for the PoF to apply to a state, the mental states required by the principle must be present within the relevant state officials.

However, the relevant mental states of the state are not always present in its officials, or are not present in the right way.

Lefkowitz notes that, often, legislative bodies in national governments are the ones tasked with enacting international law on behalf of a state. Given, for instance, that the EU adopted Operation Triton, we might infer that the member states believe the benefits of the scheme outweigh its costs. But this inference does not go through. On the state officials account, member states only have this belief if their representatives in the EU's legislature do, and representatives may vote as they do for any number of reasons. Log-rolling and coalition building, for instance, are both well-documented phenomena within the EU (Aksoy 2012; Boekhoorn et al 2006).

Lefkowitz makes the same point using the US Congress. Legislators might vote as they do "in order to secure support for the creation of a new domestic law they want Congress to enact, or to win the good graces of a powerful committee chairman who can steer government spending to their district", or for any number of other private reasons (Lefkowitz 2011, 336). Thus, state officials may enact laws or go along with schemes they believe do not promote the well-being of the state. If legislators rarely make decisions based on calculations of costs and benefits for their state, the PoF will have little to no power to create obligations at the state level. But this is not the only problem.

A further, more significant worry has to do with coherence. Since, on the state officials account, the mental states of the state are constituted by the mental states of its officials, situations in which several individuals are empowered to make a decision for the state may lead to no coherent mental state at all. If 100 legislators vote unanimously for a bill, but for 100

different reasons, many of which are incompatible and none of which involve the idea that the benefits to the state outweigh the costs, what, on the state officials account, is the underlying intention of the state? This situation results in a piece of behavior on the part of the state, but unless there is some way to coherently aggregate the mental states of the legislators, this behavior seems to correspond to no mental state at all. In terms of our picture of agency, this situation represents a failure of the ability to process and act on representational and motivational states. The state is doing something, but it is not doing it as an agent.²⁹

In response to these problems, Lefkowitz notes that while state representatives may disagree about benefits being worth their cost or disregard these considerations altogether, they may also take them into account and agree. Thus, his view by no means rules out the possibility that states come to have obligations of fair play (Lefkowitz 2011, 337). Furthermore, given the current state of international organization, the cost of participating in existing schemes is small while the benefits are large. Therefore, most officials in most states probably believe that the benefits their states receive from international public-goods schemes are worth their cost (Lefkowitz 2011, 339-40).

While the second consideration makes the first significantly more powerful, they remain worryingly contingent. Part of Lefkowitz's goal is to deflect the claim that the PoF generates unstable obligations for states. Otherwise, the principle won't provide a solid basis for the obligations needed to justify the norms and rules of mutually beneficial interactions. On the state officials account, however, the PoF only reliably generates obligations when officials deliberately act in the best interests of the state, and when the costs of participating in existing

²⁹ Lefkowitz seems to think these problems will not arise when the attitudes of only one official are at stake. He is right that disagreement will not generate coherence problems since the one official has no one to disagree with. However, the views of the official may not map on to those of the state for precisely the kinds of reasons Lefkowitz highlights.

international schemes are low. As global interconnections become denser and the costs of contributing rise (or as very inconsiderate people rise to power), so will the incentives to free-ride. It will be precisely at this point that we will need to appeal to the principle of fairness to condemn non-contributors, but it is also precisely this point at which the state officials account gives out.

The underlying problem is that Lefkowitz allows the cost-benefit calculations of individual officials to determine what counts as a benefit for the state. This view is unsatisfying. To extend the principle of fairness to artificial agents such as states or corporations, we need an account of the well-being of a state that does not rely too heavily on thick mental states, that can handle disagreements between officials within the state, and that can handle disagreements between the mental attitudes of officials and those of the state, whatever they turn out to be.

3. Perfectionism as State Well-Being

The PoF only imposes obligations when the benefits received from a scheme outweigh the costs of doing one's part in it; these benefits and costs are best understood as positive and negative impacts on one's well-being. Theories of well-being are generally divided into three camps: hedonistic theories, preference- or desire-satisfaction theories, and objective list theories. Hedonistic theories hold that one's life goes well to the extent that pleasurable experiences outweigh painful ones over the course of one's whole life. Desire satisfaction theories can take a variety of forms, but the central thought is that one's life goes better to the extent that one's welfare-related desires are satisfied. These are plausibly construed to be idealized desires about one's own life. The process of idealization can go several ways, but its goal is to prevent misinformation, poor judgement, and the like from distorting evaluations of one's well-being.

Stipulating that the relevant desires are those concerning one's own life simply keeps one as the object of one's own well-being.³⁰

Objective list theories hold that there is a list of determinate goods constitutive of well-being and that one's life goes well to the extent that it contains these goods. Given this, hedonism turns out to be a kind of objective list theory. It holds that there is one item on the list of objective goods—pleasure. Usually, however, objective list theories place several items on the list along with pleasure. One noteworthy subset of objective list theories is perfectionism, according to which the well-being of a thing consists in perfecting its distinctive nature, whatever that turns out to be. This provides an attractive unity to the list of objective goods, though, as with all the theories mentioned here, the adequacy of perfectionism as a general theory of well-being is controversial.

In this chapter I do not attempt to resolve controversies surrounding particular theories of well-being. While I will argue that perfectionism is an attractive conception of well-being for states, I remain neutral on discussions of well-being for individuals. This neutrality rests on the idea that we can be well-being pluralists, holding that the well-being of one kind of entity might be radically different from the well-being of another kind of entity. The idea is not just that a hedonist might think that pleasure for a pig is different from pleasure for a human, but that, for instance, hedonism might be the appropriate approach to well-being for a pig, desire-satisfactionism for humans, perfectionism for plants, and so on. Our approach should be to carefully examine the entity in question and develop an account of the good for that thing in light of our observations. We should not decide ahead of time that one theory will cover everything to which the concept of well-being plausibly applies. Thus, perfectionism may or may not be an

³⁰ For an overview of hedonistic, desire-satisfaction, and objective list theories, see Arneson 2016. For further discussion of each theory, see Haybron 2016, Bykvist 2016, and Hurka 2016 respectively.

attractive account of well-being for natural agents, but this, I claim, is irrelevant to the question of whether it is an attractive account of well-being for states as collective artificial agents. With this assumption in place, the task is to see which theory of well-being best fits states.

Richard Arneson provides four tests we can use to evaluate theories of well-being. The tests are meant to appeal to uncontroversial but important goods and bads such that if a theory of well-being cannot accommodate them, or struggles to, we should be suspicious of that theory.

The first test is The Cheeseburger Test. To pass it, a theory must recognize the importance for well-being of “simple ordinary pleasures of daily life such as eating a cheeseburger (or veggie burger) and watching a colorful sunset” (Arneson 2016, 594-95). Hedonism and some objective list theories pass this test easily, but both desire-satisfactionism and perfectionism struggle. A hedonist will happily recognize the importance of ordinary pleasures as will an objective list theorist (such as Arneson) who puts simple pleasures on the list of objective goods. For a desire-satisfactionist, however, if I do not care about simple ordinary pleasures (even though I find them pleasurable), their presence in my life does not enhance my well-being. The same is true for a perfectionist who does not see these pleasures as contributing to perfecting my nature.

The other three tests are The Duck Test, The Pain Test, and The Friendship Test. According to the duck test, the satisfaction of perverse or highly eccentric desires never contributes to well-being, not even a little. Squashing a duck with a large rock just because I desire to, or just because it gives me pleasure, does not contribute to my well-being. Desire or pleasure, according to the test, does not give me weak or defeasible reasons to squash the duck: they give me *no* reason to. According to the pain test, terrible pain detracts from my well-being, even if I desire it or it contributes to the perfection of my nature. Finally, the friendship test says

that lasting and meaningful friendships are valuable even if, overall, they are a pain, undesired, or unhelpful for perfection (Arneson 2016, 595-97). These tests are aimed at developing an account of well-being for human persons. Our goal, however, is to develop at least the basic contour of a theory of well-being for states as collective artificial agents. How do the three types of theory, when applied to states, fare according to the tests?

Hedonism is a non-starter as a theory of well-being for states. As I noted earlier, states may have thin mental states that allow them to represent the world, motivate them to change the world, and synthesize this information as the basis of action. However, I take it for granted that states do not have subjective experiences such as pleasure and pain. So, the problem for hedonism is not so much that it fails the tests, but that it cannot apply to states in the first place. If this were the case for all plausible theories of well-being, we would have to conclude that well-being is a concept that does not apply to states.

Desire satisfaction theories can be applied to states to give an account of their well-being, but here the duck test seems decisive. No matter what kind of entity wants to do it, squashing a duck does not contribute to its welfare. Even sophisticated desire theories that appeal to idealized self-directed desire fail the duck test. Just as a version of myself fully aware of the facts and thinking clearly might still want to squash a duck, so might a state. What's more, as Arneson notes, becoming fully aware of the facts related to one of my desires may cause me to abandon it when I shouldn't. What matters is not simply whether I would revise my desires in light of new information, but whether that information "somehow suggests reasons that render the aim unworthy or not valuable" (Arneson 2016, 604). This leaves object list theories with perfectionism as the primary candidate among them.³¹

³¹ Arneson's Bare Objective List theory could also provide an account of the well-being of states. As he notes, however, perfectionism has the attraction of providing a unified account of why the list includes the items it does.

Perfectionism's failure to address the pleasure and pain tests is not relevant here since states do not have subjective experiences. As for the duck test, pursuing aims that are needlessly cruel or highly eccentric (such as collecting belly-button lint or counting blades of grass) seems more likely to corrupt a state's nature rather than perfect it. As for friendship, it is not obvious that states need, or can even have, friends. Still, if a state can have friends, making friends of other countries, for instance, will often be valuable for the security and prosperity of its citizens. As I will argue, this suggests that it should contribute to the well-being of the state. So, at least provisionally, perfectionism is an acceptable account of well-being for states.

But this conclusion may be too quick (even having set aside issues related to friendship). Arneson argues that a thing's nature (its essential properties) is not a reliable guide to its well-being. Many things have morally negative natural capacities, and the essential capacities of a thing are not always stable. Arneson imagines a cat receiving "fancy genetic therapy" that boosts its cognitive capacities to the point that it can learn quantum physics (Arneson 2016, 593). This, he says, would be good for the cat even though it lies outside its previous natural capacities. We might think this is no problem. Now that the cat's capacities have shifted, so has an appropriate conception of its perfection. The problem, however, is about whether we should give the genetic therapy to the cat in the first place. Arneson's claim is, other things equal, we should because it would be good for the cat, but this cannot be explained by a perfectionism focused on perfecting a thing's existing capacities. Knowing that it would be good for me to perfect the nature I have now does not tell me how to change my nature if I have the opportunity.

He argues that, in the case of persons, that account is unsatisfactory. However, if, as I argue, some more unified theory is satisfactory for states, it should be preferred to a bare objective list. I therefore set it aside, though, if necessary, the view I defend could rely on it for its account of state well-being.

We might modify the perfectionist claim slightly by arguing that perfection lies not just in realizing one's nature, but also in augmenting it whenever possible. (But which augmentations should one choose if there are incompatible options?) This approach, if successful, would address the cat objection, but not the negative nature objection. If a cat can be cruel, it should not work to perfect its capacity for cruelty. Here, the problem for perfectionism is that it must operate with a value-neutral conception of a thing's nature. If we say that "what is good for a being is developing and exercising its valuable natural capacities, your claim presupposes some further unspecified account of what is valuable (good) and what is not" (Arneson 2016, 594). Cats, humans, and states have capacities for violence and cruelty, and developing these capacities seems not just morally undesirable, but bad for those who develop them.

The question that is relevant here is how these considerations apply to artificial agents. In the case of humans, animals, and other natural objects the problems raised above have two sources. One is that the nature of a natural object is not inherent in it and, if it is cognitively sophisticated, is open to interpretation and reinterpretation by that very object (as in Sartre's slogan, "existence precedes essence" (Sartre 2007, 20)). The other source is that the well-being of humans and animals is morally important. Their well-being may not be the only thing that is morally important, but we cannot simply destroy them or deprive them of well-being when it suits us. Their well-being puts constraints on how they may be treated by morally responsible agents. But, I propose, neither of these considerations applies to artefacts. Their nature is externally given, and their well-being is not intrinsically valuable. Thus, perfectionism's problems as a theory of well-being for natural agents do not arise when the theory is applied to states.

The essential difference between artefacts and natural objects is that natural objects are simply found while artefacts are created. Thus, the essential nature of an artefact is given to it from the outside and is open to revision from the outside. A sword may begin its career as an instrument of war, at which point perfecting its nature requires sharpness and durability. Later, it may be retired from combat and used in knighting ceremonies. This changes its nature. Now, perfection involves aesthetic and symbolic qualities such as elegance and cultural significance. What its nature is depends on how it is being used as an artefact. We might think this change in use causes the war sword to go out of existence and the knighting sword to come into existence. This strikes me as implausible, but it makes no difference morally since, either way, the well-being of the sword has merely instrumental value.³² Almost always, artefacts have no intrinsic moral value, so there is no problem, morally speaking, with destroying them or maintaining them at a very low level of well-being. If they are worth preserving, it is because of their instrumental value, not because of the intrinsic value of their well-being.

I believe this analysis captures well our attitudes toward artefacts. Consider a bomb. If we ask what makes for a good bomb, the first thing we need to know is what it is for. Bombs can be put to a variety of uses but let us imagine that the purpose of this bomb is to cause large, destructive explosions. For bombs of this type, nuclear weapons are the most perfect devices we have yet devised. But this does not at all impede the moral judgement that the well-being of nuclear weapons has negative moral value in the sense that it would be better if none existed, and if it could be arranged, all existing devices should be disarmed and destroyed.

³² Some artefacts may have certain kinds of intrinsic value. Great works of art, for instance, may have intrinsic aesthetic value. Also, the creation of any material artefact will involve the use or transformation of some natural resources which may have and retain intrinsic value. So, my claim is not that artefacts never have any intrinsic value. Rather, it is that their well-being as artefacts is determined by the function assigned to them. There is then a further question about whether, other things equal, it is good to promote the well-being of the artefact. And unless an artefact falls into some special category (e.g. it is a great work of art), its well-being has no intrinsic value.

Applied to states, these arguments imply that the concept of well-being does apply to states and that its content for a particular state is determined by how it is seen and used by those authorized to determine its nature. This raises the question of who is so authorized. If it is a collection of individuals, the problems of the state officials account threaten to reappear. But this need not be the case. The nature of some artefacts is fixed by social convention similar to the way in which the meaning of words is fixed. A word's meaning in a community is fixed by the generally accepted pattern of usage within that community. The same goes for many artefacts. If I take a kitchen knife and use it to turn screws, I will be told, correctly, that I am using the knife incorrectly. The point is not merely that other tools would turn the screws more effectively, but that that is not what a kitchen knife is *for*.

What if I made the object and insist that it is a screwdriver made to be indistinguishable from a kitchen knife? Intuitions may vary about whose determination is authoritative, but there is clearly a strong pull toward the idea that if an artefact is generally recognized by the surrounding community as having the nature of a kitchen knife—its purpose is to be an effective cutting instrument, particularly as a part of food preparation—then, considered merely as an artefact, that is the nature it has. Applied to states, there may be disagreement over the nature of a particular state, and there certainly will be disagreement over the shape its nature ought to take, but so long as there is enough agreement, the state will have a definite nature. And since we observe the necessary agreement for words and many other artefacts, there is no reason to doubt the same agreement is present for states.

This account of state well-being is importantly different from the state officials account of state mentality. The kind of agreement in question is a broad-strokes agreement about the general nature of the state, not specific agreement about the merits or demerits of particular

decisions. The question is not whether anyone in fact makes a particular cost-benefit calculation for a state at a particular time, but rather, whether there is enough agreement in the surrounding society about the nature of the state that such a calculation could in principle be made.

What these arguments show, or at least plausibly suggest, is that we can make cost-benefit analyses for states that allow us to apply the PoF. On a perfectionist account of the well-being of collective artificial agents, the PoF can generate obligations for states. What remains to be seen is what is the exact content of state well-being, when that well-being is morally important, and when the duties and obligations of the state generate obligations and duties for those within it to bring it about that the state fulfills them.

4. Alignment of Individual and State Well-Being

The answers to all three questions are closely related. As I have argued, the well-being of states, as artefacts, is flexible and open to revision. There may be some minimal set of functions a state must perform to survive or to qualify as a state rather than as some other kind of institution, but what this set of functions includes is not important here. For the most part, the functions of a state are determined by its officials and the people it governs. Whether perfecting its nature according to those functions is valuable depends on how those perfections affect the intrinsic values it interacts with. Perfecting the ability of a military dictatorship to suppress political opponents is not valuable. If perfecting this ability counts as perfecting the state, this shows that the nature of the state ought to change.

An especially important dimension of intrinsic value affected by the actions of a state is the well-being of its citizens and other persons globally (the well-being of animals and maybe of plants and ecosystems should also be considered). For the well-being of the state to matter

morally, its well-being must be aligned with the intrinsic values it interacts with. A state ought to see it as its function to protect and promote the well-being of its citizens without compromising the well-being of others if not actively promoting it. When the well-being of a state is aligned in this way with the well-being of its residents and others, promoting its welfare can have moral weight, as can imposing unfair burdens on it.

I do not take alignment to be a sufficient, or even a necessary condition for state legitimacy.³³ A small number of people could freely and voluntarily consent to be governed by a Hobbesian sovereign who they know will oppress and impoverish them without providing any significant benefits of any kind. This arrangement is plausibly legitimate given the free and voluntary consent given to the sovereign. Alternatively, a warlord might overthrow local legitimate political structures but go on to rule benevolently. In the first case, the authority of the sovereign is (plausibly) legitimate even though the well-being of the state is not aligned with the well-being of its citizens. In the second case, the authority of the warlord is (plausibly) illegitimate even though the well-being of the state is aligned with the well-being of its citizens.

Rather than legitimacy, alignment tells us about the moral importance of the well-being of the state. In the first case, causing damage to the well-being of the state is no reason not to act (and may even be a reason *to* act). In the second, damaging the well-being of the state might be quite a strong reason not to act depending on how thoroughly the well-being of the state is aligned with the well-being of its residents. Legitimacy may provide its own moral reasons not to interfere with the operations of a state, but these reasons are independent of, and may sometimes conflict with, reasons provided by alignment.

³³ Legitimacy here refers to moral legitimacy—having the moral right to rule—not popular legitimacy—being generally perceived as having the moral right to rule.

It should by now be clear how this account of state well-being answers the negative nature and unstable nature objections to perfectionism from the previous section. If part of a state's function is to oppress its own people, for instance, this function can be suppressed or eliminated. The well-being of its citizens (and anyone else affected by its actions) provides a set of constraints on appropriate state action even when those constraints negatively affect the well-being of the state. This conclusion is a truism, but it provides an answer to both objections. It allows us to change or eliminate negative aspects of states, and specifies "some further...account of what is valuable (good) and what is not" (Arneson 2016, 594). This further account tells us how to change and augment the nature of the state when the opportunity presents itself. Thus, the objections to perfectionism canvassed in the previous section do not apply to perfectionism as an account of state well-being, even if they apply to it as an account of individual well-being.

Let us return to the PoF by recalling the EU example in which Italy begins Operation Mare Nostrum, bears costs, and benefits other EU member states. According to what has been argued so far, do the other states have an obligation under the PoF to bear some of Italy's costs? Let us suppose that the answer is yes—they have an obligation to do their part. Nothing follows automatically, however, about what these states should do since states, as collective artificial agents, have no independent agency. They must act through the agency of their officials and citizens, so the question is whether they should act in the ways necessary to cause the state to fulfill its obligations. To determine this, we must assess the moral weight of the state's obligations.

In the EU example, the costs to Italy of running the operation are presumably passed on to citizens in the form of taxes. To that extent, the well-being of the state is aligned with the well-being of its citizens. Thus, defraying these costs carries some moral weight. If the benefits

provided by the scheme are similarly aligned with the well-being of all affected, officials and citizens ought to make their states fulfill their duties. But suppose this is not the case. Imagine that part of the nature of the states involved is to protect their borders, but that this protection provides no significant benefits to citizens or anyone else. Further, imagine that migrants picked up by the operation are placed indefinitely in refugee camps in which living conditions are comparable to or worse than the ones they left behind, and certainly worse than those they would face if they were accepted into European society. Benefits accrue to states, but not to anyone else since, here, the well-being of the states is not at all aligned with the well-being of those affected. In this case, promoting the well-being of states carries no moral weight, so there is no reason for those within the benefitting states to act to make their states fulfill their fairness obligations to each other.

But what of the burdens borne by the Italians? These carry moral weight, so perhaps they provide a reason to contribute to the scheme. I believe this is mistaken. Suppose I work for a factory making drones for an unjust war. The company that owns the factory has a contractual obligation to provide a shipment of drones in two weeks. If I slack off, the shipment will be late (so the company will fail to meet its obligation) and others in the factory will have to do more because I do less. Is it wrong for me to slack off? What is clear is that it is not wrong for me to cause the company to fail to meet its obligation which, by hypothesis, has negative moral value. The fact that other workers do more because I do less is not good, but it is not obvious what this requires of me. I might, for instance, use part of my paycheck to compensate them for the extra work they do, but since the well-being of the company is not aligned with my (or, let's assume, anyone's) well-being, I have no particular reason to defray costs by contributing to the scheme.

In the EU example, however, other Europeans have *no* reason to help Italy bear the costs of its scheme. Germany qua artificial agent may benefit from Italy's scheme, but no German citizen does. Thus, the benefit Germany receives has no moral weight, and so neither does its obligation. There is nothing tying German citizens to the scheme. If my brother promises to help his friend build a tree house, and he needs my help to do it, I have no special reason to if I don't want to even though I could significantly reduce my brother's burdens.

It is also worth noting that the alignment does not imply that benefits received by a state will be similar in kind to the benefits that devolve to its citizens. Suppose the sole function of several neighboring states is to enrich themselves. To this end, they coordinate to establish a set of trade laws enrich all the states involved. The scheme is successfully implemented and state revenues increase. Taxes and prices for individual citizens, however, do not go down since the states use the scheme to enrich themselves, not their citizens (they hoard their new wealth in vaults). However, markets within each state become more stable and predictable which promotes the well-being of everyone affected. If state S receives the benefits of this organization, do its citizens and officials have an obligation to act to cause the state to fulfill its obligation under the scheme? In this case, the answer is yes. Even though enriching the states does not enrich their citizens, their interests are still aligned since benefits to the state translate to benefits to citizens. No doubt the well-being of the states could be better aligned with the well-being of their citizens, but even partial alignment gives some moral weight to the benefits the states receive and ties those within them to the scheme from which they (indirectly) derive benefits.

Are there cases in which benefits to the state are aligned with benefits to the people, but costs are not? These cases may be harder to come by, but I believe they can be constructed. Imagine that Italy's navy is run entirely by volunteers and donations so that it costs taxpayers

nothing. It imposes no burdens on volunteers or donors they are unwilling to pay, and they may withdraw their time or resources at any time for any reason. Also, assume for the sake of argument that no Italians have any moral obligation to contribute to the navy. Operation Mare Nostrum might put a heavy strain on the finances and personnel of Italy's navy, and this may be a significant cost to the state, but it is not a cost to the people, and no people benefit. The costs of the scheme for Italy do not have any instrumental moral weight, and therefore have no moral weight at all. Given this, there is nothing morally objectionable about other states, as artificial agents, free-riding on Italy's scheme. No costs of any moral significance are borne by anyone to operate the scheme. Unfair costs may be imposed on Italy by other states, but since those costs don't matter morally, they provide no reason for anyone to cause their state to fulfill its obligation to contribute.

A final comment is in order about liberty and self-determination. We began with the idea that liberty is very important. For our discussion here, the relevant question has been how liberty affects well-being. As we have seen, states are not intrinsically morally important. This implies that their well-being is not intrinsically important, and it implies the same about their liberty and self-determination. However, the ability of a state to be self-determining, especially through democratic institutions, may be very important to the well-being of its citizens. To the extent that this is the case, the liberty of a state gains instrumental moral importance. What this means is that, if the PoF places restrictions on the ability of a state to be self-determining, this may not be a strike against the PoF and the obligations it generates. What matters is how these restrictions affect the well-being of those whose well-being is intrinsically important. And if the PoF can sometimes restrict the liberty and self-determination of individuals whose well-being does have intrinsic moral value, then it certainly can restrict the liberty and self-determination of states.

5. Causal Efficacy of State Obligations

One might react to the preceding arguments by suspecting that states as agents play no causal role in the obligations of their citizens. Perhaps I am right that states can, in a formal sense, have fair-play obligations, but these obligations are like ghosts. Even if they exist, they have no substance. In non-metaphorical terms, any fairness obligation a state acquires that translates to obligations for officials and citizens is the result of a situation that would produce the same obligations for the same individuals whatever we say about states. Imagine that the establishment of the EU Schengen area was initiated at the state level and produces a variety of public-good benefits for both states and citizens. In this case, states have an obligation to contribute their fair share to maintaining the area, and citizens have an obligation to cause their states to fulfill this obligation. But to whom do citizens owe this obligation? We might think that the obligation is owed to everyone in the Schengen area who contributes to her state doing its part since this is the mechanism by which all the individuals affected receive benefits and contribute costs. If this is the case, however, benefits to the states are an idle wheel. Their well-being is simply irrelevant so long as benefits accrue to individual persons within the Schengen area.

The objection has considerable force in overdetermination cases such as this, though even here state obligations may not be totally inert. For instance, imagine that the benefits provided by the Schengen area are spread unevenly. The German state benefits a lot though its people benefit only a little, while the Spanish state benefits only a little though its people benefit a lot. Proportionality requires Germany the state to bear greater costs than Spain the state, and requires the German people to bear lesser costs than the Spanish people. There is at least an open question about how complications of this sort should be handled, but I set them aside to focus on clearer

cases in which fairness obligations at the state level play an important causal role in creating obligations for individuals.

Imagine that a state passes an international treaty into law because it benefits the state. Some segment of the population is benefitted by this treaty but let us imagine that it is a relatively small segment. This ensures that, for nearly everyone, the effects of the treaty won't directly create PoF obligations to comply with its terms. Assuming that the treaty is enacted, compliance with it becomes a matter of domestic law. Depending on our theory of political obligation, this may be enough to generate a duty to comply with the terms of the treaty. There may also be purely domestic benefits to living under a state that reliably fulfills its obligations, for instance. In these ways, then, the ability of states to acquire PoF obligations can play a causal role in the obligations of residents.

More generally, the well-being of a state can be aligned with intrinsic moral values other than individual well-being (assuming such values exist). Suppose that China institutes a massive and costly clean-energy program that will significantly mitigate the severity of climate change over the next 100 years and will make large contributions to adaptation and compensation as needs arise. By the time these benefits start affecting the lives of individual persons, most people now living will be dead. However, since states can survive indefinitely, we can imagine that the scheme begins benefitting other states immediately (assuming their well-being involves promoting and protecting the well-being of their populations over time). So, even though future generations do not yet exist, the fact that they, as a group, will be better off means that the program promotes states' well-being now. We might also imagine that protecting the natural environment has intrinsic moral value, and that this value is aligned with the well-being of many states.

In either case, state well-being inherits the moral significance of the value its well-being is aligned with, even if this value does not involve the well-being of current citizens. In this case, the benefits and burdens associated with a public-goods scheme operating at the state level are morally important. This gives those within each state moral reasons to ensure their state fulfills its obligations, even if this means bearing extra costs themselves. In this way, the PoF as applied to states can play a primary causal role in the obligations of ordinary citizens.

Our initial question was whether states and other collective artificial agents could bear obligations under the PoF. The first task was to determine whether these agents have the mental capacities necessary for the principle to apply. I argued that thin representational states are sufficient for recognizing moral reasons as moral reasons and acting on them for that reason, and that this is sufficient for treating states as responsible moral agents. The next task was to determine whether states can receive benefits. This would require applying a theory of well-being to states, and I have argued that a particularly attractive candidate is perfectionism. The nature of a state is determined by the community that constitutes it, and its well-being has no intrinsic moral value though it may have instrumental value if the functions that constitute its nature are aligned with intrinsic moral values. Finally, I have argued that, on this account, the ability of a state to bear obligations under the PoF can causally contribute to significant obligations for its citizens that cannot be accounted for through independent applications of the PoF. Thus, not only can artificial agents bear fairness obligations, but this ability can play an important role in helping us understand our obligations to each other, domestically and internationally.

Chapter 2, in full or in part, is currently being prepared for publication of the material. Finley, Aaron. The dissertation author was the primary author of this material.

References

- Aksoy, Deniz. 2012. "Institutional Arrangements and Logrolling: Evidence from the European Union." *American Journal of Political Science*, 56 (3): 538-552.
- Arneson, Richard. 2016. "Does Fairness Require a Multidimensional Approach?" In M. D. Adler, and M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*, 588-614. New York: Oxford University Press.
- Boekhoorn, Vincent, Adrian M. A. Van Deemen, and Madeleine O. and Hosli. 2006. "A Cooperative Approach to Decision-Making in the European Union." In R. Thomson, F. N. Stokman, C. H. Achen, and T. König (Eds.), *The European Union Decides: Testing Theories of European Decision-Making*, 178-210. Cambridge: Cambridge University Press.
- Bykvist, Krister. 2016. "Preference-Based Views of Well-Being." In M. D. Adler, and M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*, 321-346. New York: Oxford University Press.
- Haybron, Daniel M. 2016. "Mental State Approaches to Well-Being." In M. D. Adler, and M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*, 347-378. New York: Oxford University Press.
- Hurka, Thomas. 2016. "Objective Goods." In M. D. Adler, and M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*, 379-402. New York: Oxford University Press.
- Lefkowitz, David. 2011. "The Principle of Fairness and States' Duty to Obey International Law." *Canadian Journal of Law and Jurisprudence*, 24 (2): 327-346.
- List, Christian., and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. New York: Oxford University Press.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Cambridge: Blackwell.
- Sartre, Jean-Paul. 2007. *Existentialism Is a Humanism*. J. Kulka (Ed.), and C. Macomber (Trans.). New Haven: Yale University Press.
- Simmons, A. John. 1979. *Moral Principles and Political Obligations*. Princeton: Princeton University Press.

Chapter 3

Fairness Without Cooperation

Introduction

Consent theories of political obligation explain the duty to obey the law by arguing that the governed, through some voluntary act or omission, have granted their governing institutions the authority to govern them. These theories have a number of attractive features, one of which is the clear link they establish between consenting individuals and the political institutions that apply to them. If Canadians have consented to the rule of the Canadian government, then they are morally obliged to obey it—they have a special obligation owed exclusively to Canada. However, consent theories are commonly, and I think rightly, thought to have a fatal flaw. While they do link the consenting to their governing institutions, they link *only* the consenting—and actual members of political communities rarely consent to being governed.³⁴ The principle of fairness (which I will often refer to simply as ‘the principle’) takes up the spirit of consent theories by defending political obligation as a type of special obligation, but makes fair reciprocity the core concept rather than consent. Specifically, the principle is built around a simple and appealing no-free-riding intuition. If others are doing their part in a scheme that produces benefits for all, it is unfair for me to receive those benefits while refusing to do my part. Applied to political obligation, when others do their part by paying taxes, obeying the law, serving on juries, and so on, I owe it to them to do my part too whether I have consented to being governed or not.

³⁴ For sustained arguments against explicit and implicit consent theories, see Simmons (1979, ch.3,4) and Huemer (2013, ch.2,3). Regarding hypothetical contract theories, Dworkin offers a succinct criticism: “a hypothetical contract is not simply a pale form of an actual contract; it is no contract at all” (Dworkin 1973, 501).

A. John Simmons has argued that the principle of fairness faces a fatal flaw, parallel to that faced by consent theories, because of its reliance on cooperation. According to Rawls's canonical statement of the principle,

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. (Rawls 1999, 96, my emphasis)³⁵

Simmons argues that Rawls's 'right to acquiescence' relies on a thick notion of cooperation. For him, genuine cooperation requires that

Conscious sacrifice for the common good... [and that] others make deliberate sacrifices in support of mutually beneficial goals, while *relying* on others to do the same—and while having reasonable grounds for such reliance. . . [on the basis of] sharing the same goals and understandings. Indeed, others act in such contexts only because they think they have reasonable grounds for believing that. . . most of us will freely make the same sacrifices. (Simmons 2001, 40, original emphasis)

He goes on to say that “it is these features of the strongly cooperative setting that make us feel that it would be wrong to fail to do our parts” (Simmons 2001, 41 original emphasis).

Large political communities, however, do not exhibit these qualities. Of course, most people usually obey the law. But, Simmons argues, this obedience is most often motivated by fear of punishment, habit, blind reverences for the law, or the recognition that the requirements of the law are also requirements of morality. Notably, none of these motives count as cooperative. And if they are not cooperative, failing to obey the law does not “constitute *taking*

³⁵ Each piece of the principle has been debated in the literature. I note here in passing that one of the most important debates is about whether the principle gives rise to obligations only for those who voluntarily accept the benefits of a scheme, or if mere receipt of benefits is enough. This debate is generally framed as a standoff between voluntarists and non-voluntarists. A. John Simmons's 1979 chapter and his later 2001 chapter have in many ways set the agenda on this and other debates about the principle of fairness. His work, as far as I know, is unsurpassed as a defense of the voluntarist position and has attracted notable responses from theorists such as Arneson (1982, 2013); Klosko (2004); Dagger (2000). Also, it is worth noting that Rawls, in the sentence before the one quoted above, appeals explicitly to the notion of voluntary acceptance of benefits which may well place him in the voluntarist camp, at least in *A Theory of Justice*.

advantage of anyone” (Simmons 2001, 42; original emphasis). Just as consent theories fail to establish a duty to obey the law because (almost) no one in fact consents to being governed, the principle of fairness fails because (almost) no one genuinely cooperates.

Simmons’s argument assumes that mutual good-will is at the core of cooperation and therefore at the core of fair reciprocity. Reciprocity is owed when several people knowingly rely on each other to make sacrifices toward a shared conception of the common good. I argue that this is not the case. Much of my argument is aimed at establishing that our no-free-riding intuitions are responsive to cases in which someone acts toward the common good, but does so unknowingly or begrudgingly. I argue that the principle of fairness is not grounded in a mutual understanding of the common good, but rather in transactional sensibilities that take respect for productive labor as their fundamental value. Very roughly, if I benefit from the efforts of others without compensating them, I can count as unfairly free-riding on their efforts even if they bear no cooperative intentions toward me. Thus, rather than attempting to argue that political communities are cooperative,³⁶ I argue that the cooperation condition should be removed from the principle altogether.³⁷

It is worth noting that rejecting the cooperation condition comes with significant apparent costs. One of the central tasks of the principle of fairness is to prevent our no-free-riding intuition from running amok. As Nozick correctly observed, “you may not decide to give me something, for example a book, and then grab money from me to pay for it, even if I have nothing better to spend the money on” (Nozick 1974, 93). The principle must identify which

³⁶ For a representative argument, see Song 2012.

³⁷ The only other arguments against the cooperation condition I am aware of are presented by Cullity (1995); Lefkowitz (2011); Armstrong (2016).

benefits trigger reciprocity obligations when received, and under what circumstances. On standard interpretations, cooperation plays a crucial role in determining those circumstances.

The cooperation condition operates through participants' intentions and can help determine the scope of the principle in at least two important ways.³⁸ For one, it prevents mere positive externalities from generating obligations willy-nilly.³⁹ According to the condition, positive externalities can generate obligations only if they are accompanied by appropriately cooperative intentions. Since most positive externalities are (presumably) not accompanied by intentions of this sort, the principle of fairness is unlikely to generate obligations when they are received.

The cooperation condition can easily explain why both unintended beneficiaries and those intended not to be beneficiaries incur no fairness obligations under a scheme even when they receive its benefits. Those outside the scheme who benefit are not intended beneficiaries. Since they are therefore not the objects of cooperative intentions, no obligations arise for them. This explains, for instance, why the obligations associated with national defense and the rule of law stop at national borders even when their benefits spread farther. Thus, the cooperation condition can play an important role in preventing the principle of fairness from overgeneralizing.

The argument will proceed in four stages. In section 1, I argue that there are cases in which cooperative intentions are absent, but free-riding still seems possible.⁴⁰ In section 2, I

³⁸ The claim is not that the intentions of benefactors must refer specifically to each individual for whom obligations arise. It is enough, for instance, if they have a generalized intention to benefit certain groups such as 'conationals' or 'all on whom the benefit happens to fall'.

³⁹ A positive externality, in the economic sense, arises when some of the benefits produced by an economic exchange are not captured by the parties to the exchange. For a negative externality, just replace the term 'benefits' with 'costs' (cf. Cooter & Ulen 2016; R. L. Miller 2014). Here, however, we can think of positive externalities more informally as the benefits one's actions produce for others for which they do not pay.

⁴⁰ One might worry that because my definition of a scheme (introduced on page 4) derives from Simmons's definition of cooperation, my discussions of schemes are automatically discussions of *cooperative* schemes. As we

defend the claim that the principle of fairness is grounded in transactional relationships. In section 3, I develop a Scanlonian interpretation of transactional relationships by casting them as legitimate reciprocation-expectations that cannot be reasonably rejected. In section 4, I argue that a transactional understanding of the principle can adequately prevent it from overgeneralizing, and in section 5, I conclude.

1. Achilles Among the Farmers

My argument centers on two examples that track a distinction between two groups within a public-goods scheme: those who produce the public good (producers), and those who merely receive it (beneficiaries). For instance, if a neighborhood engineers' association maintains an air purification pylon that cleans the otherwise harmfully dirty air, those in the engineers' association are the producers of the good of clean air from which all in the neighborhood benefit. Everyone else in the neighborhood is a beneficiary. Because they benefit, the beneficiaries owe their fair share of the costs of the scheme in which the producers work to generate the good that benefits them all. Using this distinction, I will argue in two stages against the cooperation requirement. First, I will show that beneficiaries can free-ride on a scheme even when they have no cooperative intentions toward each other. Next, I will show that free-riding is possible when beneficiaries and producers bear no cooperative intentions toward each other. Thus, however we

have seen, however, cooperation essentially involves the intention that others benefit from my sacrifices; that is the point of Simmons's talk of "conscious sacrifice for the common good" (Simmons 2001, 40). Supporting mutually beneficial goals, however, need not have the same connotation. Mutually beneficial goals frequently are not common goals. I might sell you metal scrap knowing that you will use it to make sculptures I find repulsive. And you might know I will use the money on an expensive sushi dinner, something you regard as wasteful and off-putting. We might vehemently oppose each other's goals while acknowledging that our joint activity (the exchange) is mutually beneficial. Since schemes, as defined, only require that goals be mutually beneficial, they do not automatically involve cooperative intentions.

carve up a particular scheme, we can be confident that cooperative intentions are not necessary for the principle of fairness to apply and generate obligations.⁴¹

The basic scenario proceeds as follows. Contra Homer, Achilles survives the Trojan war, but instead of returning to his homeland, he abandons Greece and settles in a small valley to become a farmer. The valley is inhabited by other farmers and has only one entrance through a narrow pass. As it happens, Achilles settles very near the entrance. Shortly after he settles, the valley begins to be menaced by barbarians. One night, Achilles sees the flicker of torches in the pass. He quickly dons his armor and weapons and proceeds to defeat the barbarians in the narrow valley entrance. The barbarians are great warriors indeed since, if they fought in the open, Achilles would be defeated. But because he fights them only two or three at a time, he is able to drive them back. Achilles is painfully wounded whenever he fights the barbarians, but, because they attack infrequently, he has enough time between attacks to recover.

There are several features of the basic case worth highlighting. For one, Achilles bears very heavy costs in the form of his periodic injuries. The other farmers can't share this cost because Achilles is the only warrior capable of standing up to the barbarians. If any of the other farmers were present, they would just get in Achilles' way and would most likely be killed. But despite the costs, Achilles has sufficient prudential motivation to defend the valley entrance regardless of the actions of the other farmers. If Achilles doesn't defend the entrance, he (and the other farmers) will die. The fact that Achilles will die if he doesn't defend the entrance is enough to motivate him to defend it, and this is known to everyone.⁴² The other farmers, aside from

⁴¹ See Boran (2006) for a discussion of the circumstances under which production of a public good can generate fairness obligations. She argues that the intentions of producers matter in the sense that accidentally produced benefits do not give rise to fairness obligations.

⁴² Thus, on my view, fairness obligations can be triggered when producers like Achilles bear any costs, not just when they bear net costs. See Boran (2006, 105) for an articulation of the view that obligations only arise when producers bear net costs.

being useless as fighters, are all similarly situated relative both to each other and to Achilles. They are all in good health, have comparable levels of wealth, are on a par in terms of social status, and so forth. The point here is not that costs must be high for the principle of fairness to kick in. Instead, proponents of the cooperation condition hold that cooperative intentions are necessary for the principle of fairness to produce obligations, so without them, no obligations can arise even when costs are high.

With this basic scenario in place, we can move on to the first variation, which is designed to show that the principle of fairness can generate obligations within a scheme even if beneficiaries have no cooperative intentions toward each other.

Imagine that there are only two farmers in the valley when Achilles arrives. Each lives alone, and while neither has ever discovered the other, Achilles has discovered both. After he begins defending the entrance, he visits each farmer separately and informs him of his efforts and the costs he bears to defend the valley. He proposes that they institute a fair cost-sharing scheme (defense for farm-goods, perhaps). After all, they all benefit equally,⁴³ and Achilles believes it would be unfair for him to bear all the costs alone. He is glad that his efforts help each of the farmers, but he wants each of them to do their part in keeping them all safe (so he has cooperative intentions toward each of the farmers). Each farmer finds Achilles' case compelling and agrees that he has an obligation to participate in the scheme. However, neither farmer knows that the other farmer also participates in the scheme.⁴⁴ Thus, neither intends his contribution to benefit the other (nor would they form such intentions if they discovered each other). Does this

⁴³ Klosko (2004) argues that our intuitions in cases like these derive from the nature of the good provided. The provision of presumptive goods (or Rawlsian primary goods) can generate fairness obligations because one needs them for any reasonable life-plan. I argue in the next section that it is not a good's importance that brings the principle to bear, but the good's nature as legitimately transactable.

⁴⁴ To avoid the possibility that Achilles initiates two schemes, we may stipulate that Achilles would not be able to maintain the scheme without contributions from both farmers.

mean that no fair play obligations arise at all for the farmers? This hardly seems likely. Clearly what matters in this case is the relationship between Achilles and each farmer. Of course, I have stipulated cooperative intentions between Achilles and the farmers, so the next case aims to show that our fair play intuitions also respond to cases in which producers and beneficiaries carry no cooperative intentions at all.

In the second variation of the case, the two independent farmers are replaced by a village of farmers who all know each other. But now Achilles does not intend that his efforts produce benefits for the other farmers. In fact, he wishes he could leave the other farmers to the barbarians and announces as much in town. Even so, he says, it is not fair that he bears the very high costs he does without any kind of reciprocation. Like any standard public-goods scenario, the good is being produced in an ongoing manner, the good is a genuine benefit, *ex ante* price negotiations are impossible, and Achilles' division of benefits and burdens is (let's assume) fair.⁴⁵ Given all this, is there any reason to think that Achilles' lack of cooperative intentions undermines any obligation the farmers might have to contribute to his scheme?⁴⁶

One might attempt to establish their importance by appealing to the bare intuition that, morally speaking, intentions usually matter and that on these grounds Achilles is not entitled to compensation given the absence of any intention to help the farmers. An intentional killing is much worse than an accidental killing, for instance, even if the actions that produce the two deaths are identical. This kind of observation may generate a presumption that intentions always matter morally unless we are shown otherwise. However, one who holds this view could still

⁴⁵ That is, given each person's preference-calibrated benefit level, everyone bears a proportional share of the costs. Those costs are substantially less than the corresponding benefits, and no one pays more than she would have agreed to pay had price negotiations been possible *ex ante*.

⁴⁶ I have set aside for section 5 the worry that the principle of fairness without the cooperation condition will generate too many obligations.

think that there is often a plurality of moral considerations that determine what we owe each other, and that, sometimes, intentions play no role. This is the view I defend here. I am happy to say that intentions are often morally important, so the burden of proof falls on me to show why cooperative intentions are not necessary for obligations to arise under the principle of fairness.

2. Fair Play as a Transactional Relationship

To illustrate the claim that cooperative intentions are not necessary for the principle of fairness to generate obligations, consider the following analogy. Suppose that Achilles provides the good of protection to the other farmers, but in doing so he not only lacks cooperative intentions, he possesses uncooperative intentions. Proponents of the cooperation condition will say that these ill-willed intentions are enough to prevent the principle of fairness from producing obligations in the farmers.

But consider another case. I walk up to a vendor running a hotdog cart and in deciding whether or not to order a hotdog, it becomes clear that the vendor only wants the worst for me. He says he hopes my buying the hotdog is reinforcing unhealthy eating habits, he insults my taste in food, and is generally rude and mean-spirited. Unlike a public-goods scenario, I can just walk away without engaging in an economic transaction—that is, an exchange of goods or services—but the fact that the vendor possesses uncooperative intentions toward me does not make it permissible for me to get a hotdog and then leave without paying, even if we are in the state of nature. My claim is that, if I am not allowed to free-ride on the productive efforts of the hotdog vendor by taking a hotdog without paying, then the farmers are not allowed to consume the good of protection from Achilles without fair reciprocation.

The two cases are relevantly similar because they involve transactional relationships that embody the same core features. In an economic exchange, such as the hotdog case, once the

transactional relationship is established, what matters for transaction-related obligations are the mutually agreeable terms of the transaction and each party's living up to those terms. These two conditions are, in David Shoemaker's terms, the "relationship-defining features" of the transaction which each party must uphold (Shoemaker 2011, 623). Established relationships governed by the principle of fairness are characterized by the same relationship-defining features as relationships centered on transactions in the market. Free-riding just is failing to live up to the mutually agreeable terms of a fair cost-sharing scheme. And when failure to contribute counts as free-riding, participating counts as reciprocation for the receipt of benefits. However, the analogy needs to be examined more closely. Transactional norms only have binding force when exchanges are initiated in the right way and in the case of market transactions, 'the right way' is through voluntary agreement. So, for the analogy to go through in a non-voluntarist framework, public-goods schemes must have mechanisms that do the same work as voluntary agreement does in a market setting.

To better understand market transactions, we can split them into three parts: part one tells us which relationships are transactional, part two tells us how transactional relationships are initiated, and part three tells us what the core features of transactional relationships are once they are established.⁴⁷ I will refer to these, respectively, as the *categorization*, *initiation*, and *core* aspect of transactional relationships. Using this terminology, I have already claimed that the core aspects of market- and public-goods-based transactional relationships are identical because both are characterized by an obligation to reciprocate for goods and services rendered according to mutually agreeable terms. However, categorization and initiation in public-goods schemes do not parallel market-transactional relationships so closely.

⁴⁷ See Phillips (1997); Phillips & Reichart (2000) for an approach to the principle that is transactional in nature, but that interprets transactions as voluntaristic exchanges grounded in consent.

Normally, the categorization and initiation aspects of market relationships are very closely related. To categorize the relationships that could be market-transactional, we need only identify parties who propose an exchange of goods or services. Not all such proposals are legitimate, of course, because not all goods are legitimately transactable (for instance, it is immoral to buy and sell people, and it is illegal to buy and sell cocaine). For a good to be minimally transactable (regardless of legitimacy), we must recognize it, when confronted with it in a context, as a good that can be bought or sold.

A market-transactional relationship is initiated when another party voluntarily bargains over or accepts a transactional proposal. Proposals need not be made or accepted explicitly, and a transactional relationship might be only one of several relationships two of agents share. Much more could be said, but, as an initial gloss, categorization and initiation can be characterized for familiar sorts of market transactions as follows: for categorization, the relevant kinds of relationships are those initiated by proposing an exchange of goods or services, and for initiation, a particular market-transactional relationship is established when a proposed exchange is voluntarily accepted or bargained over.

To vindicate an analogy between public-goods schemes and standard market transactions, I must show that their initiation and categorization aspects function similarly. In a public-goods scheme, the function of initiation is performed by receipt of benefits. But, one of the central problems with providing public goods is that bargaining over costs or rejecting the good is unfeasible. However, we can still use bargaining as an ideal model by appealing to the price to which one would freely agree if *ex ante* price negotiations were possible. Assuming the right kind of scheme has been established (that is, that categorization conditions are already satisfied), what one owes within a scheme is what one would have agreed to pay had *ex ante* negotiations

been feasible, or a share of the costs of the scheme in proportion to the benefits one receives, whichever is lower. This is what one may, in fairness, be expected to contribute to the scheme, and therefore constitutes the terms on which the transactional relationship is initiated.⁴⁸ For categorization, too, we may take our cue from familiar market transactions and lean on the notion of transactable goods. A transactable good is any good or service that we recognize as a candidate for buying or selling in a given context. As noted above, there are both moral and legal constraints on the class of legitimately transactable goods, but there are conventional constraints as well. Only some goods are transactable, and only in certain contexts.⁴⁹ Humans, for instance, are not transactable, but labor usually is. And since we get fair-play intuitions about Achilles' defense efforts, his labor in that case is, plausibly, transactable.

There are two important points to highlight about the gloss of transactional relationships I have just presented. First, it captures the idea that public-goods schemes are similar enough to market transactions to be governed by the same fundamental relational norms. The core aspects of the two kinds of relationships are identical, and the categorization and initiation aspects are analogous. Second, this gloss captures the intuition that what matters for transactability is the *nature* of goods and services up for exchange, not the attitudes or intentions of the people who produce them.

So, the gloss reveals a crucial point—what matters morally for obligations of fairness is not cooperative intentions, but the nature of, and circumstances surrounding, the good provided. For public goods, the work of categorization is done by providing transactable public goods in

⁴⁸ For parallel arguments in context of third-party beneficiaries of wrongdoing, see Butt (2007, 2014).

⁴⁹ See, for instance, Walzer's list of blocked exchanges (Walzer 1983, 100-103), and Andre (1992) for an attempt to organize Walzer's list by kind. Note also that some goods may be legitimately transactable for only some kinds of transactions. We think, for instance, that it is okay for the public to pay judges indirectly to give rulings, but not okay for defendants to pay them directly. The labor performed by judges is therefore legitimately transactable, but only when the transaction is of the right kind—in this case public rather than private.

mutually beneficial and contextually appropriate ways (assuming that *ex ante* negotiations are not feasible). The work of initiation is carried out by the double upper limit of hypothetical agreement and a division of costs in proportion to benefits received. The core aspect of both market-oriented and public-goods-oriented transactional relationships is the same. One owes reciprocation for goods and services rendered within properly initiated transactional relationships. And, as I have argued, public-goods provision is rightly conceptualized by analogy to market transactions.⁵⁰

3. Transactional Relationships as Legitimate Expectations of Reciprocation

I have argued that market transactions for private goods and transactional relationships based on public-goods schemes are morally on a par, but we might wonder why *any* transactional relationship is morally important. To explain why obligations that arise in these relationships have normative force in general, I appeal to a version of Scanlonian contractualism oriented toward legitimate expectations.

Adapting Thomas Scanlon's characterization of moral wrongness, we can say that failing to live up to the transactional expectations of another is wrong if, under the circumstances, no one could reasonably reject the expectation as a component of informed, unforced transactions.⁵¹ The need for 'informed' and 'unforced' should be clear; they are present to prevent expectations from legitimizing fraudulent or coercive transactions. The term 'reasonable' here can be interpreted much as Scanlon does. He says, "the intended force of the qualification 'reasonably,' ...is to exclude rejections that would be unreasonable *given* the aim of finding

⁵⁰ Cf. Arneson (1982) for an articulation of a similar view on Nozickian grounds. Arneson argues that Nozick (1974) doesn't have the theoretical resource to reject the principle of fairness in the way he claims to in *ASU*.

⁵¹ The wording of this principle is adapted from, and very closely mirrors, Scanlon's principle describing the wrongness of an action in general (Scanlon 2013, 597).

principles which could be the basis of informed, unforced general agreement” (Scanlon 2013, 597, original emphasis). However, I allow the term ‘reasonably’ to be morally loaded in a way Scanlon does not. He argues that his contractualist formula can serve as the basis for morality itself. Clearly, my transactional-expectations norm is doing no such work. Thus, for instance, I take it for granted that one may not reasonably expect someone to fulfill a demand she knows to be immoral on independent grounds.

These sorts of constraints aside, the kinds of expectations that plausibly have binding obligatory force in transactional relationships are those that arise from the relationship-defining features of the core aspect of the transaction (mutually beneficial terms of exchange in the absence of fraudulent intentions, and the upholding of those terms). This basic formula may be structured by more finely grained local practices, but something like the normative requirement that, other things equal, we live up to the reasonable expectations of others, gives rise to obligations in transactional relationships even in the absence of cooperative intentions.⁵²

What is relevant about the transactional relationships described above—what makes a relationship transactional in the appropriate sense for the principle of fairness—just is that the parties to the relationship can reasonably expect fair reciprocation for goods and services rendered. In the case of market transactions, fair reciprocation is defined by actual agreements, and in the case of public-goods provision, it is defined by hypothetical agreements. So, in both contexts we can say that the agreement in the transactional setting is what gives rise to obligations, and those obligations cannot be undermined unless the expectations defining the transactional relationship itself are undermined by one of the parties to it. For instance,

⁵² Not everyone will think this kind of deeper justification is necessary—at least not in the public-goods case. Arneson, for instance, feels no need to articulate a “deep explanation of what might justify the principle [of fairness]. To my mind its attraction is simple and lies on the surface” Arneson (2013, 134).

‘agreements’ made under duress do not bind because no genuine agreement can be reached. Similarly, one owes nothing to a fraudulent⁵³ transactor, even if she bore costs in deliberately rendering a good or service that was not agreed to, because she does not uphold her end of the agreement.

On this analysis, what is distinctive about public-goods scenarios is that they, under the right conditions, describe situations in which one may non-voluntarily come to be a party to a transactional relationship. Our no-free-riding intuitions are sensitive to these situations, and the principle of fairness aims to systematize the conditions under which they arise. What the Achilles cases show is that in transactional relationships involving public goods, just as in those involving private goods, what we owe each other *merely qua transaction* does not depend on the presence of cooperative intentions.

The intuition to which I appeal, then, is that, in the case in which Achilles unavoidably bears large costs to produce benefits for all, the other farmers owe Achilles their participation in a fair cost-sharing scheme even when neither Achilles nor the farmers have cooperative intentions toward each other. My explanation is that Achilles, by producing the good and proposing the scheme, establishes a setting in which he and the farmers are parties to a transactional relationship. Because of the unusual features of public-goods schemes, it is not permissible for the farmers to opt out of the relationship, and because of the relationship-defining features of economic transactions, what matters are not cooperative intentions, but mere receipt of benefits. We now have good reason to think that cooperation is unnecessary between beneficiaries, and between beneficiaries and producers.

⁵³ Fraud occurs when a party to a transaction intends at some point before the termination of the transactional relationship to violate the terms of the transaction, and then acts on the intention. Perhaps mere fraudulent intentions are sufficient to break a transactional agreement, but either way, no cooperative intentions need be present for obligations to arise. Rather, specifically fraudulent intentions need only be *absent*.

4. Overgeneralization

I now return to the worry I raised in the introduction that, without cooperative intentions, the principle of fairness overgeneralizes. I do this by addressing three types of cases in which the principle might seem to generate too many obligations. The first worry I address is that the principle of fairness, as I have developed it, applies to positive externalities in general, rather than just to public goods under relatively narrow circumstances. For the other two types of cases I consider, I take my cue from A. John Simmons, who claims that a version of the principle like mine will apply, implausibly, to very small benefits and unintended benefaction where harm was intended (Simmons 2001, 39). I will address each concern in turn.

Positive externalities can, on the face of it, look problematic for this version of the principle. Consider the following case: the US government, operating as normal, produces public goods that benefit not only US residents, but those living just across the Canadian border as well. One person in particular receives benefits not provided by the Canadian government that are worth significantly more than her fair share of what she would pay in US taxes for the goods if she were a US resident. Furthermore, she would be willing to pay at least that amount had it been possible to negotiate the price of provision. Does she owe taxes to the US government for the benefits she receives?

The correct answer is, it depends on the context. In the world as it is, when one lives and operates entirely within the country in which one is a citizen, any expectation that one should pay taxes to any government but the one that rules in one's territory can be reasonably rejected.⁵⁴ So long as no such expectations are in place, no public-goods schemes will be established across national borders in the apparently problematic way I have just described. Establishing

⁵⁴ Tax laws are complex, and I do not mean my generalization to cover all situations. The rule of thumb I am aiming at is, unless one operates in a country in some way, one is not expected to pay taxes to that country.

expectations that can't be reasonably rejected are crucial for establishing reasonable demands within transactional relationships. Without those expectations, no obligations arise.

What if expectations were established? The US government could, for instance, make a public announcement and set up a website through which Canadians could make their payments. The website might even include easy-to-use tables and formulas that would reliably indicate how much one owes (if anything) depending on one's situation. I am willing to accept that, in this case, if the benefits received by those along the Canadian border are large enough, then they would indeed owe something to the US.

While this may be a bullet to bite, it is not as large as it may seem. For one, the same problem arises for versions of the principle that rely on cooperative intentions. Suppose, for instance, that US residents are generally good-natured and intend that the public goods they produce through their government benefit all who receive them so long as recipients contribute their fair share. In this case, they would be glad to discover that Canadians are benefitting from US public-goods schemes and might even be willing to take on additional costs to ensure that they continue to benefit. According to standard formulations of the principle, obligations arise for Canadians in this case too, and, in fact, they may arise even more readily than they do on the version of the principle I have defended. If mere intentions do significant work in setting the boundaries of the obligations produced by public-goods schemes, a productive group of good-natured individuals could produce many obligations for those around them in a way that seems problematic. What this shows is that my formulation of the principle handles positive externalities no worse than standard formulations.

More importantly, we must keep proportionality considerations in mind; benefits must be worth their costs for obligations to arise. First, notice that in a very extreme case, in which

Canada is in complete anarchy and those near the US border receive massive benefits, there seems to be no bullet to bite at all in saying that those in Canada who benefit owe something to the US. But as conditions in Canada improve, the contributions owed to the US shrink, and they probably shrink quite rapidly as basic goods, such as security and the rule of law, are established. Furthermore, it would be costly both diplomatically and administratively to establish and maintain the systems needed to share costs across borders. If these costs plus the preexisting costs of the scheme outweigh the benefits of the new cost-sharing systems, then no obligations arise. So, it is in precisely those cases in which it seems intuitively problematic for obligations to arise internationally that the principle of fairness, as I have construed it, is least likely to produce them. The theory does not rule out the possibility of international fairness obligations, but I do not think it is a desideratum of a theory of fair play that they be ruled out in principle.⁵⁵

We may now move on to the first of Simmons's two problem cases: does the principle of fairness, without cooperative intentions, give rise to too many obligations by creating obligations to reciprocate when very small benefits are received? Consider a case of a very small, purely positive benefit. Suppose I am a wonderful singer, and, just to benefit my neighbors, I practice my singing in the village green instead of in my home where no one would hear me. Those who hear me benefit, but the benefit is minuscule. There are several reasons obligations do not arise in a case like this. For one, in this case, any feasible payment that could be made by those who benefit would be comparable to the benefits received. For this reason, the benefits will not clearly be worth their cost, so no obligations arise. Furthermore, it would be unreasonable to think that those who benefit would have agreed to pay anything to listen to me practice my

⁵⁵ It is not clear we should be hostile to the possibility of the principle of fairness generating significant obligations between members of different states. For instance, see Miller (2009); O'Neill (2005) for discussions of the ways in which political obligations do not respect national boundaries.

singing. For this reason also, no obligations arise. Finally, since, as we may assume, I am singing just for fun, or simply because I want to, the good of singing is costless for me to produce. A fair share of no costs is no costs, so no obligations arise.⁵⁶

Benefits and costs come in degrees, so at some point, benefits will be large enough to produce obligations, so long as there are attendant costs to share. However, these benefits will have to be fairly substantial before significant obligations begin to arise. So, this version of the principle does not create obligations in response to very small benefits.⁵⁷

Lastly, we might wonder what to make of cases in which a number of persons institute a scheme intended to produce a public bad, but “act on *misinformation* sufficient to accidentally produce a [public good]” (Simmons 2001, 39, original emphasis). Could obligations of fairness arise to contribute to a scheme of this sort? In this case, the details matter. For one, the producers, presumably, have fraudulent intentions. Thus, the setting is not properly transactional, so no obligations will arise on the part of beneficiaries toward producers. However, depending on the kind of uptake the scheme receives, other participants might come to owe each other their continued support of the scheme. But, here again, the situation of the ill-willed producers matters. Their ill-willed attempts to produce harm might count as a threat to those who end up benefiting. For this reason, any obligations that might otherwise arise may well be displaced by the need to defend against the threat posed by the producers. If they could be deposed and replaced by other people who would maintain the scheme in good faith, then, if any obligations arise, it would be an obligation to replace the current producers. If that course of action is not

⁵⁶ We might also note that the good could be made excludable. I could practice at my house and charge admission. Thus, *ex ante* price negotiations are feasible, so no obligations arise. This can be easily amended by stipulating that I have no option but to practice on the village green.

⁵⁷ However, I do not claim, like Klosko (2004), that obligations only arise when the benefits received consist of goods necessary for a minimally decent life. (See Koltonski (2016) for a detailed critique of Klosko.) On my view, any public good can produce fairness obligations so long as it is legitimately transactable.

feasible, however, then it is not the case that the scheme is up and running in the appropriate sense. What's more, given that the will of the producers is to restructure the scheme such that the goods it produces are replaced by bads, expecting beneficiaries to contribute amounts to expecting them to support a scheme deliberately intended to hurt them. Rejecting this expectation certainly would not be unreasonable "*given* the aim of finding principles which could be the basis of informed, unforced general agreement" (Scanlon 2013, 597, original emphasis).

5. Conclusion

I began by highlighting a potential parallel between consent theories of political obligation and theories based on the principle of fairness. Consent theories argue that political authority is not legitimate unless the governed give their consent to be governed. However, it seems unlikely that more than a few do. Similarly, Simmons has argued that the principle of fairness requires cooperative intentions, but that very few members of modern political societies have them. I have responded to this objection by arguing that cooperation is not an integral part of the principle of fairness. Mutuality is not a necessary aspect of the principle, but rather respect for the productive labor of others. Even though Simmons's argument may go through—it may be the case that very few people have cooperative intentions about the public goods produced by their political institutions—the parallel does not. Consent is clearly essential to consent theories, but, as I have argued, cooperation is not essential to the principle of fairness. Thus, I have defended a version of the principle of fairness that avoids Simmons's objection and places the principle on firmer ground for supporting political obligation.

The point has not been to argue that cooperative intentions are a bad or that the principle is incompatible with them. For all I have argued, they might reinforce, or sometimes even create,

fairness obligations. Rather, what I have argued is that cooperative intentions are not *necessary* for fairness obligations. So, while it may or may not be true that most people lack cooperative intentions toward those who contribute to the functioning of political institutions, it is plausible that, at least in reasonably just societies, most people benefit from their operation. Basic goods such as the rule of law, national security, and infrastructure, are valuable goods (though questions remain about how the value of these and other goods should be measured, which goods count, and who in society receives a net benefit from them given the total package of government activities). Thus, the possibility remains open that, even in relatively just societies, some do not receive a net benefit from their political institutions and therefore have no obligation to obey the law under the principle of fairness (though there may often be other reasons to obey the law).

Political authority is sometimes construed as *general*—it applies to everyone (or nearly everyone) in a territory—and *content-independent*—the state is often entitled to enforce laws even when they are unjust or otherwise unhelpful.⁵⁸ Whether the principle of fairness obligates enough people to obey the law for political authority to count as general is an open empirical question that will be importantly informed by our theory of distributive justice. Under the principle of fairness, political obligation is also sensitive to the *content* of the law. As political institutions produce fewer net benefits, the duty to obey them shrinks and, at some point, disappears entirely. The cooperation condition can bolster these obligations so long as most members of society intend for others to benefit. Obligations apply generally to the objects of cooperative intentions, and the presence of cooperative intentions might weigh in the scales

⁵⁸ See Huemer (2013, 12-13) and Simmons (1979, chapter 2) for more extensive accounts of the nature of political authority.

alongside efficiency. Thus, by removing the cooperation condition, my version of the principle might seem to weaken the generality and content-independence of political authority.

This may be so, but I see it as a strength of the argument I have presented, not a weakness. The principle's emphasis on benefits helpfully highlights the importance of the conditions of people's lives and the extent to which the actual activities of governments help or hurt them. Harmful and wasteful policies ought to be replaced, whatever the intent behind them. In these cases, it is not the thought that counts. To the extent that removing the cooperation condition can help remove support from unjust and inefficient schemes, so much the better.

Chapter 3, in full or in part, is currently being prepared for publication of the material. Finley, Aaron. The dissertation author was the primary author of this material.

References

- Andre, Judith. 1992. "Blocked Exchanges: A Taxonomy." *Ethics*, 103 (1): 29-47.
- Armstrong, Christopher. 2016. "Fairness, Free-Riding and Rainforest Protection." *Political Theory*, 44 (1): 106–130.
- Arneson, Richard. 1982. "The Principle of Fairness and Free-Rider Problems." *Ethics*, 92 (4): 616–633.
- Arneson, Richard. 2013. "Paternalism and the Principle of Fairness." In C. Coons and M. Weber (Eds.), *Paternalism: Theory and Practice*, 134–156. Cambridge: Cambridge University Press.
- Boran, Idil. 2006. "Benefits, Intentions, and the Principle of Fairness." *Canadian Journal of Philosophy*, 36 (1): 95–115.
- Butt, Daniel. 2007. "On Benefiting from Injustice." *Canadian Journal of Philosophy*, 37 (1): 129–152.
- Butt, Daniel. 2014. "'A Doctrine Quite New and Altogether Untenable': Defending the Beneficiary Pays Principle." *Journal of Applied Philosophy*, 31 (4): 336–348.
- Cooter, Robert., and Thomas Ulen. 2016. *Law & Economics, 6th Edition*. Berkley Law Books.
- Cullity, Garrett. 1995. "Moral Free Riding." *Philosophy and Public Affairs*, 24 (1): 3–34.

- Dagger, Richard. 2000. "Membership, Fair Play, and Political Obligation." *Political Studies*, 48 (1): 104–117.
- Dworkin, Ronald. 1973. "The Original Position." *The University of Chicago Law Review*, 40 (3): 500–533.
- Huemer, Michael. 2013. *The Problem of Political Authority*. Palgrave MacMillian.
- Klosko, George. 2004. *The Principle of Fairness and Political Obligation*. Rowman & Littlefield.
- Koltonski, Daniel. 2016. "The Principle of Fairness, Political Duties, and the Benefits Proviso Mistake." *Journal of Moral Philosophy*, 13 (3): 265–293.
- Lefkowitz, David. 2011. "The Principle of Fairness and States' Duty to Obey International Law." *Canadian Journal of Law and Jurisprudence*, 24 (2): 327–346.
- Miller, David. 2009. "Justice and Boundaries." *Politics, Philosophy and Economics*, 8 (3): 291–309.
- Miller, Roger LeRoy. 2014. *Economics Today, 17th Edition*. Arlington: Pearson.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Cambridge: Blackwell.
- O'Neill, Onora. 2005. "Justice and Boundaries." In *Political Restructuring in Europe*, 65–84. Routledge.
- Phillips, Robert A. 1997. "Stakeholder Theory and a Principle of Fairness." *Business Ethics Quarterly*, 7 (1): 51–66.
- Phillips, Robert A., and Joel Reichart. 2000. "The Environment as a Stakeholder? A Fairness-Based Approach." *Journal of Business Ethics*, 23 (2): 185–197.
- Rawls, John. 1999. *A Theory of Justice, Revised Edition*. Cambridge: Harvard University Press.
- Scanlon, Thomas. 2013. "Contractualism and Utilitarianism." In R. Shafer-Landau (Ed.), *Ethical theory: An anthology, 2nd Edition*, 593–607. Wiley-Blackwell.
- Shoemaker, David. 2011. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics*, 121 (3): 602–632.
- Simmons, A. John. 1979. "The Principle of Fair Play." In *Moral Principles and Political Obligations*, 101–142. Princeton University Press.

Simmons, A. John. 2001. "Fair Play and Political Obligation: Twenty Years Later. In *Justification and Legitimacy: Essays on Rights and Obligations*, 27–42. Cambridge University Press.

Walzer, Michael. 1983. *Spheres of Justice*. Basic Books.

Song, Edward. 2012. "Acceptance, Fairness, and Political Obligation." *Legal Theory*, 18: 209-229.

Chapter 4

Slack-Taking and Burden-Dumping: Fair Cost-Sharing in Duties to Rescue

Introduction

Globally, millions of individuals need rescue from disease, natural disaster, poverty, and violence. If everyone in a position to perform rescues did her fair share, no one's share would be large. But when some individuals fail to do their part, how much slack must others take up? Singer, Unger, and others have argued that we have very stringent duties to do more when others do less.⁵⁹ Many, including L. J. Cohen, Liam Murphy, and David Miller, have argued in response that principles requiring one to take up slack are objectionably unfair. These principles, they argue, demand too much from conscientious individuals by requiring them to do not only their share, but also the shares of those who neglect to do their part. Even worse, the principles seem to let the morally negligent off the hook by making their burdens the responsibility of others.

I agree that contributing more than one's fair share to a rescue effort is unfair but disagree that principles are the source of the unfairness. Instead, by shirking their responsibilities, non-contributors unfairly dump part of the burdens they should have borne onto others. Thus, the conduct of burden-dumpers, far from being permissible, constitutes a double wrong—they wrong those they fail to rescue, and they wrong those on whom their burdens fall. On this approach, those who do their part have an obligation to take up at least some slack, and burden-dumpers remain responsible for failing to do their part.

Importantly, I do not defend the act-consequentialist position that those who do their part must take up *all* the slack left by others. The view I defend here is consistent with deontological

⁵⁹ Singer (2007); Unger (1996). Singer defends both a strong and a weak principle of beneficence and argues that both are very demanding. In this chapter, I set aside questions about the demandingness of our duties. Thus, I will usually mention the strong version of Singer's principle, not because I take it to be most plausible, but because it most sharply highlights the contours of the debate.

views that posit a duty to perform rescues so long as they are not too costly. So long as my fair share of the burdens under full compliance is less than the maximum this duty could require of me, others failing to do their share increases my burdens. This chapter does not address the quantity of slack that must be taken up, arguing instead that those who leave slack treat slack-takers unfairly.

Because my central focus is on fairness rather than demandingness, I aim to describe cases in which our natural duty to rescue is clear. Singer's famous drowning child example will therefore be central. Sadly, however, children drown in the real world as well. In 2013, a vessel left Libya carrying around 500 migrants. En route, it caught fire and sank off the coast of Lampedusa, a small Italian island in the Mediterranean. Upwards of 300 of those on board died. The incident attracted international attention and Italy used its navy to begin a search and rescue program called Operation Mare Nostrum that is credited with rescuing some 200,000 people during the year it operated. However, due to the cost of the program, Italy appealed to the rest of the EU for help. In response, Frontex, the EU's border and coast guard agency, was tasked with replacing Italy's program with a new one—Operation Triton. Triton has received criticism for focusing primarily on border control rather than search and rescue. This has left a serious humanitarian crisis in the Mediterranean as thousands of people die or go missing each year attempting to cross from North Africa to Europe.

The crisis in the Mediterranean provides a vivid example of the kind of large-scale, ongoing rescue efforts we face. Italy and other EU member states recognized a duty to perform at least some rescues, and they seemingly recognized an obligation toward Italy to share the burdens of performing those rescues. Italy claimed it would be unfair for it to bear the burden of performing all the rescues alone, and others in the EU apparently agreed. It is this intuitive

connection between natural duties and fairness obligations that I develop here. I argue that when duties to rescue require someone to do more than her fair share (the amount she would have to do under full compliance), she is being treated unfairly by the people who fail to do their part. This argument draws on features of the literature on group causation and moral responsibility. In particular, I combine Alvin Goldman's vector theory of causation with David Brink's and Dana Nelkin's fair-opportunity theory of responsibility. I argue that non-contributors treat contributors unfairly by failing to do their part when a) the failure derives from a blameworthy lack of responsiveness to features of a situation (such as drowning children or overly burdened rescuers) that give one moral reasons to act, and b) the failure imposes burdens by leaving slack that contributors must take up.

I lay the groundwork for addressing questions of responsibility under partial compliance in section 1 by articulating an account of the content of our duty to rescue. In section 2, I address a puzzle related to the following question: on whom exactly do burdens fall when non-contributors fail to do their part? After elucidating the puzzle, I defend a solution in the context of non-discretionary duties to rescue.⁶⁰ In section 3, I expand on the arguments developed in sections 1 and 2 and show that they cover discretionary duties to rescue. More specifically, I argue that when one culpably fails to do one's part, one is implicated in generating the burdens one's failure, together with the similar failures of others, could reduce. This means that when one's duty to rescue is discretionary, one treats all contributors unfairly by failing to do one's

⁶⁰ A non-discretionary duty is a duty with only one means of fulfillment. If I promise to do X, I don't keep my promise unless I do X. A discretionary duty is one that I may fulfill as I see fit. If I have a duty to help the badly off, I could work at a local homeless shelter, donate to Oxfam, dig wells, and so on. The contrast between the two types of duties is not deep. A non-discretionary duty is just a discretionary duty with only one fulfillment option. However, the distinction is useful because non-discretionary duties are an important subclass of duties and are easier to analyze than discretionary duties.

Some theorists identify non-discretionary duties with Kantian imperfect duties. Murphy briefly discusses this view in Murphy (2000, 71-72), and Ignieski (2006) analyzes duties to aid in terms of Kantian perfect and imperfect duties. For further discussion of imperfect duties as such, see Baron (1987) and Hope (2014).

part. In section 4, I consider some objections, and in section 5, I consider further applications of the theory focusing on voting and climate change.

1. The No-Burden-Dumping Intuition

In his landmark paper, “Famine, Affluence, and Morality”, Peter Singer argues that we are obligated to use our resources to rescue those dying from lack of food, shelter, and medical care up to the point of marginal utility—the point at which further sacrifice would make us worse off than those we are helping (Singer 2007, 507).⁶¹ He then considers a series of objections, one of which concerns fairness: if each affluent person contributed to ending this kind of suffering, no one would have to donate more than a few dollars (Pogge 2011). We are all morally required to contribute, so isn’t it grossly unfair that I, the conscientious person, must donate to the point of marginal utility simply because others aren’t doing their part? In response, Singer says it is unfortunate that others aren’t contributing, but that doesn’t change the fact that we have a duty to rescue as many as we can even if others aren’t. For Singer, the decisive consideration is that “by giving more than £5 [what I would give under full compliance] I will prevent more suffering” (Singer 2007, 507).

I will assume that we do have a duty to rescue, and that this duty does require us to do more when others do less. However, I set aside the question of how much more we are required to do.⁶² What I want to draw out is the intuition that those who do not contribute to the rescue

⁶¹ This is the strong version of Singer’s argument. The weak version says only that we must give until doing so would force us to sacrifice something of moral significance. Since both the weak and strong versions are very demanding, one can raise the fairness objection to both.

⁶² For a book-length discussion of demandingness in the context of taking up slack, see Murphy (2000). There (and in Murphy 1993), he argues that we are never required to do more than we would be if everyone were doing her part, even in drowning child cases. I do not have the space to take up his arguments here, but for concise and forceful replies see Keith Horton (2004, 2011). Horton argues that one’s objection to doing more than one’s fair share becomes increasingly weighty the more slack one must take up. Horton suggests that this unfairness, in addition to the extra costs one bears, weighs against one’s duty to take up slack past a certain level of sacrifice.

effort wrong not only those they fail to rescue, but also those who take up the slack. I will refer to those who don't do their part, and thereby leave more work for others, as *burden-dumpers*.

The claim I defend is not about how much can be demanded of us, but about who or what is at fault when that demand is unfair. Cohen, Murphy, and Miller all argue that the principle making the demand is the source of the unfairness, but I argue that it is not. Rather, the unfairness originates in the people who neglect their duties.

Throughout the discussion we must carefully separate the wrong of neglecting one's duty to rescue from the wrong of burden-dumping. Consider a variation of Singer's drowning child example. I and another person are near a pond in which two children are drowning. The other person and I could easily save one child each. However, I see that if I do nothing, the other person will be able to save both children, though just barely. I decide to do nothing, and the other person saves both children. I will say that the child has a *deontic* complaint against me because she had a duty to her that I failed to fulfill. In general, deontic complaints arise when one fails to fulfill an individual duty to another agent that is not generated by a maldistribution of resources.⁶³ I will say that the other rescuer has a *fairness* complaint against me because she had to do more than her fair share because of my culpable failure to do my part.⁶⁴ This case raises questions that I will briefly address before turning to a puzzle about collective burden-dumping.

However, I am inclined to agree with Anja Karnein (2014) that this kind of unfairness does not weigh against one's duty to take up slack. Instead, it should be counted against non-contributors in determining, for example, what kind of compensation they might owe to those who took up their slack.

⁶³ Michael Ridge (2010) suggests that when I fail to perform a rescue, and no one takes up my slack, I treat the un-rescued person unfairly. This claim is sensible since my failure produces a maldistribution of burdens. Because I failed to bear the burden of performing the rescue, the person in need of rescue must bear the consequences of remaining un-rescued. This observation raises questions about the proper delineation of duties by kind that I do not have space to address here.

⁶⁴ Perhaps the other rescuer has a right against me that I not impose undue burdens on her. Even so, the complaint is about fairness because it concerns a maldistribution of resources. I forced her to use her resources to perform a rescue when it should have been my resources that were expended.

Imagine that the other rescuer and I are positioned such that it is initially unclear which child I should save. When I perform no rescues, have I wronged both children, or neither? Does either of the children have a deontic complaint against me? A detailed discussion is beyond the scope of this chapter, so here I suggest that my duty only becomes particular once it is clear which child the other rescuer is going to save. We might think that neither child has a right against me that I save *her*, though, plausibly, each has a right that I “save as many of them as [I] could without unreasonable risk to [myself]” (Feinberg 1984, 61).⁶⁵ Whatever rights the children might have held against me, it is clear that I had a duty to rescue at least one of them and that I wrong both children by simply ignoring it.

What if the burden I shirk is too heavy for the other person to carry, but she can still shoulder some of it? Suppose there are three children drowning and that I could save two children as easily as the other swimmer could save one. Other things equal, my duty is to save two, while the duty of the other is to save one. However, I save none. Through tremendous exertion the other rescuer saves two children, but the third child still drowns. Clearly the third child has been wronged, but by whom? Given the language of burden-dumping, one might think that I dumped my duty to save my two children onto the other rescuer, so that only she wrongs the third child by failing to rescue her. If nothing else, the-ought-implies-can principle entails that the second rescuer isn’t obligated to save all three children. But we can be more precise about each rescuer’s obligations: each must perform as many rescues as she can given a) her relevant abilities (for example, how strong a swimmer she is), b) the scope of the need, c) the

⁶⁵ For Feinberg, the sign that this right exists is our sense of moral indignation at potential rescuers when they do nothing (1984, 64). See Agnafors (2017) for a further defense of disjunctive rights and Wolterstorff 2008, chapter 11 for a general discussion of correlativity between rights and duties.

total costs she can reasonably be required to bear,⁶⁶ and d) the portion of the need others can be expected to satisfy.

Condition d) becomes important in cases in which burdens must be fairly distributed. If fairness does not demand that others share my burdens, I can't expect them to contribute anything and conditions a)-c) determine what I am obligated to do. However, once fairness comes into the picture, we might worry that d) gives non-contributors a free pass to dump their burdens so long as others are willing to take up the slack. If I am in the presence of several conscientious individuals, I might know that if I do nothing, all the drowning children will be rescued. Given this, d) seems to let me off the hook. Because they can be expected to do everything, I have no obligation to do anything. What's more, as L. J. Cohen worries, if everyone knows there is a duty to take up slack left by non-contributors, even those who are inclined to contribute "could legitimately infer that, if he failed to do so, those with tenderer consciences than himself would make good the deficiency. So any temptation that he might have to withhold his own contribution would be reinforced by the belief that...the ultimate outcome would be the same" (see Cohen 1981, 73-74).

This objection highlights an ambiguity in the notion of expectation employed in condition d). On the one hand, according to a fair distribution, others can be expected—in the sense of normatively required—to contribute their initial fair share. On the other, according to their actual attitudes, they can be expected—in the sense of predicted—to contribute whatever they are willing to contribute, which may be as little as nothing. Both notions of expectation are relevant here, and both generate obligations. According to fairness, one is responsible for one's initial fair share of the burdens even if one contributes nothing. But if some can be expected to contribute

⁶⁶ Even Singer's strong principle would endorse condition c). After all, I might be able to give past the point of marginal utility, but Singer thinks I am not morally required to.

less than their fair share (according to their actual attitudes), the rest of us are obligated to take up their slack. One person's unwillingness to do her part affects the scope of the need facing others—condition b)—without changing the portion of the need she is normatively required to address. In this way, those of us who contribute become responsible for the burdens of non-contributors, even though the non-contributors remain responsible for their share of the rescues.⁶⁷

What follows is that, in the case in which I, in fairness, ought to save two of three drowning children, but save none, while another rescuer does her best and saves two, only I wrong the third child. However, I also wrong the second child by failing to fulfill my duty to her. What's more, if the other rescuer saved only one child, both of us would wrong both of the other children. I obviously wrong both since I could have saved both, and the other rescuer wrongs both since she could have saved either. She is guilty of a deontic failing toward them by violating her duty to save as many as she can without unreasonable risk to herself. So, my culpable failure to rescue makes a similar culpable failure possible for the other rescuer. As the number of potential rescuers grows, there is no upper bound to the amount of morally culpable wrongdoing a single problem can produce so long as we are all duty-bound to solve it.⁶⁸

2. Collective Burden-Dumping

At this point, a puzzle might seem to arise. Suppose six children are drowning, Jones, Smith, and I are the only potential rescuers, and each of us has a non-discretionary duty to rescue the children. We are all on a par as swimmers, and each of us can save two children easily but

⁶⁷ What I say here may not fully address Cohen's worry about temptation. On one level, since condition d) doesn't let non-contributors morally off the hook when others take up their slack, no one can be tempted by the possibility of avoiding wrongdoing while also failing to do her part. But if Cohen's point is merely psychological, I have nothing to say one way or another. But see Michael Ridge (2010) for an argument against Cohen's claim about perverse incentives.

⁶⁸ See Karnein (2014) for an argument that comes to a similar conclusion on this point.

cannot save more than three. In this case, each of us ought to save two children—that is a fair distribution of rescue-related burdens. I immediately rescue two children. By the time this is done, I see that Smith and Jones intend to save no children. So, my obligation to take up slack kicks in and I save a third child after which it is too late for the other children.

Intuitively, I have a fairness complaint against Smith and Jones for imposing the burden of performing a third rescue. Both fail to contribute to the rescue effort, so both play a role in the extra burdens I bear.⁶⁹ But Jones might say, “Smith was unwilling to help, so if I had helped, you and I would have saved three children each. You were already saving three children, so *I* dumped no burdens on you.” And Smith could say the same. (Call this case *partial help*.)

Smith’s and Jones’s argument seems sensible because it appeals to an intuitively plausible characterization of what it means to play a role in someone’s burdens. According to Jones and Smith, one plays a role in another’s burdens only when one’s contribution would alone be sufficient to reduce the burdens borne by contributors. The complication in this case is that Smith and Jones impose burdens jointly, not individually. So, is there a defensible sense in which each⁷⁰ plays a role in my burdens even though neither, acting alone, could reduce them—a defensible sense in which I can still properly raise a fairness complaint against each?

Alvin Goldman defends a potential answer to this question in his analysis of the obligation to vote. Those who vote or refrain from voting almost never cast or withhold a decisive ballot. Thus, those who do not vote, or vote for a bad candidate, can run an argument parallel to Jones’s and Smith’s. Each person can say that *her* vote or abstention did not affect the outcome of the election, so *she* should not be held responsible.

⁶⁹ I use the admittedly awkward phrase ‘play a role’ to avoid using the word ‘contribute’ to refer to opposite phenomena—contributing to rescue efforts and contributing to burdens by failing to contribute to rescue efforts.

⁷⁰ By ‘each’ I mean each individually, not both of them collectively. For a discussion of similar cases in the context of collective responsibility, see Björnsson (2020).

Goldman responds to this objection by developing what he calls a vector-system analysis of causal contributions. He explains:

[A vector is a sum] computed from three kinds of forces: (1) forces that are positive in the direction of movement, (2) forces that are negative in the direction of movement, and (3) forces that are zero in the direction of movement. Finally, when thinking about the causation of a given movement, we think of each positive force as a *contributing factor* in the production of the movement, each negative force as a *counteracting*, or *resisting*, factor in the production of the movement, and each zero force as a *neutral factor* vis-a-vis the production of the movement. (Goldman 1999, 210-11 original emphasis)

Each person who casts a vote for the winning candidate is a causal contributor to—or, in my terms, plays a role in—that person’s victory. Similarly, in the case of Jones and Smith, each plays a role in the extra burdens I bear since the inaction of each is a contributing factor in them.⁷¹ But if Jones helps rescue while Smith does not, Jones’s action counts as a force in the direction of distributing burdens fairly. Thus, even though Jones’s action does not reduce the burdens I bear, his change in behavior changes the direction of his vector contribution.

The vector analysis, however, is incomplete as an account of responsibility. Suppose I arrive at the polls intending to vote for the best candidate, but as I put pen to paper an unforeseeable muscle spasm causes me to vote for the worst candidate after which a strong gust of wind blows my ballot into the counting machine. My vote is a contributing force in the direction of the bad candidate, but I am clearly not responsible for the contribution. We might similarly wonder about those who lack reliable transportation, or who are misinformed about the

⁷¹ Goldman’s vector account is best interpreted as an extension and smoothing out of J.L. Mackie’s INUS conditions for causation in “Causes and Conditions”. For instance, an INUS analysis of voting is different for even and odd numbered electorates in ways that seem to reflect theoretical machinery rather than the ethics of voting (Goldman 1999, 206-10). The vector account does not face similar technical complications. Still, if a candidate won in a landslide, why did my vote count as a causal contributor when the outcome we care about is not the scalar ‘force’ of the votes, but the binary of victory and defeat? In my view, something like Mackie’s INUS analysis is still needed to answer this question. My vote contributed to the victory because in some subset of votes for the candidate, mine was necessary for her victory. Thus, Goldman’s theory is best seen as extending or reformulating Mackie’s.

candidates,⁷² anxious in crowded places, forgetful, and so on. What we need is a systematic way to distinguish between those who are responsible for the role they play in dumping burdens, and those who are not.⁷³

David Brink and Dana Nelkin defend a reasons-responsive account of responsibility⁷⁴ on which blameworthiness requires a fair opportunity to avoid wrongdoing. This fair opportunity has three parts: a cognitive component, a volitional component, and a situational component. Briefly, the cognitive component involves the “capacity to make suitable normative discriminations, in particular, to recognize wrongdoing” (Brink 2013, 132). The volitional component involves “the capacity to regulate one’s actions in accordance with this normative knowledge [one’s recognition of right and wrong]” (Brink 2013, 132-33). Finally, the situational component involves “*external or situational* factors... [such as] *coercion* and *duress* [which] may lead the agent into wrongdoing in a way that nonetheless provides an excuse, whether full or partial” (Brink 2013, 134 original emphasis).

In the case of voting, failure to vote (or voting for someone other than the best candidate) is blameworthy when the three conditions listed above are satisfied. Cognitively, this requires, for instance, that information about the candidates’ policy stances and qualifications are readily available and intelligible. Volitionally, one must be able to vote according to one’s considered convictions rather than peer pressure, a candidate’s charisma, or other irrelevancies. Situationally, voting must not jeopardize one’s employment, expose one to undue risks, or be

⁷² See Goldman 1999, 210 for some discussion.

⁷³ Because those who vote are not required to vote more when others vote less, failing to vote does not dump burdens. I discuss the relevance of my arguments to voting in section V.

⁷⁴ See Brink’s and Nelkin’s “Fairness and the Architecture of Responsibility”, and Brink’s “Situationism, Responsibility, and Fair Opportunity”, especially section IV. Modern theories of responsibility fall into two broad categories, reasons-responsive theories and attributionist theories. I employ a reasons-responsive theory of responsibility as one I take to be plausible, though not uncontroversial. For recent defenses of attributionism, see Sher (2009) and Smith (2008). For an overview of the debate, see Talbert (2016).

otherwise inaccessible. So long as these conditions are met, and so long as one lives in a legitimate democracy, one can be blamed for failing to vote.

A similar analysis can be given for the duty to rescue, though the analysis is complicated by the fact that failing to rescue can cause two distinct wrongs—the deontic wrong to those one is duty-bound to rescue, and the fairness wrong to others involved in the rescue effort. I will consider the cognitive and volitional components first. Cognitively, duties to rescue are usually easy to understand and information about rescue efforts is widely distributed and easy to find. Volitionally, fulfilling these duties often requires no more than donating to effective organizations. Similarly, the distributional implications of partial compliance are widely understood. At some point, bearing extra burdens will strain one’s volitional capacities, but even if this excuses one from bearing the full weight of one’s obligations, one must still work as close as possible to the point of critical volitional stress. Plausibly, most individuals satisfy the cognitive and volitional requirements for responsibility in relation to their duties to rescue and their obligations to distribute the burdens of those rescues fairly. Henceforth, I set aside cognition and volition and focus on the situational component of responsibility.

Within this reasons-responsive framework, Jones might run the following argument against the claim that she is responsible for dumping burdens. First, she might acknowledge that she is responsible for failing to perform rescues, and that in some mechanical sense this failure ‘contributed’ a vector force pushing in the direction of burden-dumping. Still, she should not be held responsible for those burdens because she did not have a fair opportunity to prevent them. Everyone knew that Smith was not going to do his part, and Jones’s contribution alone could not make a difference in the burdens I bear. It is therefore unfair to hold her responsible, even partly, for dumping burdens since she had no opportunity to do otherwise. We might reply by noting

that Jones could still have done her part, in which case her vector contribution would have changed and she would no longer count as a burden-dumper. But this does not quite capture the spirit of Jones's reply. Her claim is that she could perform rescues, so she had reason to, but she could not lighten my burdens, so she had no reason to. If Jones was faced with no distribution-related reasons, I cannot blame her for taking no distribution-relation action. To evaluate this objection, it will be helpful to consider a case in which Jones is unable to act unless Smith acts.

Suppose I have been poisoned and will soon be dead if no antidote is administered. The antidote consists of two ingredients each of which is ineffective if administered alone. As it happens, Jones and Smith have one ingredient each, and both are present. Unfortunately for me, Smith refuses to give up his ingredient for morally indefensible reasons (he likes the look of its color). Jones, however, rushes to my side to do what she can for me, entreating Smith to do the same. But without Smith, Jones can't help me, and I die. Intuitively, and I think rightly, Smith is responsible for my death and Jones is not. One possible explanation is that Smith, unlike Jones, had a fair opportunity to make a difference to the outcome. Whatever Jones did, she could not prevent my death. On this line of reasoning, she did not have the relevant situational control, so she cannot be held responsible. If this were right, Jones might argue that the same line of reasoning applies in the partial help case. There too she cannot be held responsible for any 'vector contribution' she makes to dumping burdens because she did not have the situational control necessary to prevent extra burdens from falling on me.

Jones's argument that the poison and partial help cases are relevantly similar conflates reasons to change outcomes with reasons to be *willing* to change outcomes. In the poison case, Jones displays concern for my condition and attempts to convince Smith to act. This shows that she is responsive to the available moral reasons. Contrast this with a case in which neither Smith

nor Jones is willing to give up their ingredients for indefensible reasons. Now, it seems, *both* are responsible for my death even though neither, acting alone, can avert it. Neither displays any willingness to do her part which is precisely what the situation calls for. If they were appropriately responsive to the available moral reasons, each would show willingness to contribute an antidote ingredient. They would then administer the antidote and I would be saved. In the partial help case, appropriate responsiveness to the available moral reasons means performing one's share of the rescues. This is something Jones can do even if she can't reduce the burdens I bear, so she does have the situational control needed to be appropriately reasons-responsive. Thus, her failure to perform any rescues marks her as a blameworthy causal contributor in the direction of burden-dumping.⁷⁵

Given the preceding arguments, I propose the following characterization of what it means to play a role in dumping burdens: one plays a role in (makes a vector contribution to) unfair burdens borne by contributors when one's failure to be sufficiently reasons-responsive in a context of fair opportunity, together with similar failures on the part of others,⁷⁶ is sufficient to impose on contributors more than their fair share of the costs to be distributed.⁷⁷

⁷⁵ What if Jones displays willingness to distribute burdens fairly, but no independent willingness to rescue the drowning children? Do I still have a fairness complaint against her? I propose that the answer is Yes. Jones ought to respond to the full set of moral reasons available to her, and partial responsiveness does not imply partial blameworthiness. Imagine that I love slashing tires which is both *expensive* and *upsetting* for my victims, and that these are the only relevant moral reasons in the situation. If I were responsive to both reasons, I would not slash tires. But I only care about upsetting people (I am fully responsive to this reason) which alone does not outweigh my enjoyment. It seems to me that when I slash tires, I am blameworthy for upsetting my victims even though I am fully responsive to the moral reasons that their distress gives me to refrain.

⁷⁶ The number of others may be zero.

⁷⁷ Michael Ridge (2010) offers an alternative solution to collective burden-dumping cases. He argues that the burdens left by non-contributors ought to be shared among rescuers and rescuees alike. Thus, *partial help* cases will not arise since any additional contribution will reduce my burdens at least marginally. This line of reasoning is mistaken because it incorrectly classifies obligees as obligors. If those to whom obligations are owed are not responsible for bearing a share of those obligations initially, it is unclear why they would become responsible when some obligors fail to contribute.

I have argued that when one fails to do one's part in a rescue effort, one treats other rescuers unfairly, at least in non-discretionary cases. What remains to be seen is whether the arguments I have laid out extend to discretionary duties to rescue. If I only have the resources to contribute to one rescue scheme, but there are five equally good schemes to choose from, who is treated unfairly when I do nothing?

3. The Particularity Problem

So far, I have focused on rescue scenarios that, by hypothesis, impose a non-discretionary duty to rescue, which in turn means that any fairness obligations are owed to other rescuers on the scene. I have a non-discretionary duty to rescue *that child* (or these children), which means that I have a duty to help *these* people perform the rescue. However, some will reject the claim that this duty is non-discretionary. Singer, for instance, argues that because our duty to rescue doesn't take distance into account, saving a child right in front of me is morally on a par with saving a child on the other side of the world (other things equal). In the same way that I may choose which drowning children to save when I can't save them all, I may choose which rescue effort to participate in when I can only participate in some.

There is controversy over whether the duty to rescue those who are close is more stringent than the duty to rescue those who are far away, and whether I have the discretion to contribute to rescue efforts other than the most efficient one.⁷⁸ I do not attempt to address these questions here, and assume for the sake of argument that distance does not matter, and that one has at least limited discretion to choose rescue options that are not maximally efficient. That said, questions about discretion arise regardless. I might be equidistant from two drowning

⁷⁸ For representative arguments, see Ingeski (2006), Feinberg (1984), Smith (1998), and Kamm (1999).

children, each of whom has a rescue effort dedicated to her, and, of course, I am confronted with a wide range of organizations to contribute to that carry out rescues all over the world. So, even if there is controversy over the degree to which duties to rescue are discretionary, there is widespread agreement that they allow for at least some discretion.

Duties to rescue being discretionary raises a potential problem for the account of burden-dumping I have defended. The claim that, by failing to contribute, I dump burdens on other rescuers seems to require a particular rescue effort to which I am bound to contribute—I must have a reason to contribute to *that effort* in particular. If there are no particularizing reasons, then there are no particular burdens I am required to help bear, and thus no answer to the question of who is unfairly burdened when I do nothing. Burden-dumping appears incompatible with discretionary duties to rescue. Call this the *particularity problem*.⁷⁹

To flesh out the problem, consider again the case of migrants attempting to cross the Mediterranean and suppose Italy is doing all it can to rescue vessels in distress. Suppose also that France and Spain have a similar duty to rescue distressed vessels. However, the need is so great that if France or Spain (not both) does all it can, there will be no less for Italy to do. But if both helped, each would carry significantly lighter burdens than it would working alone or in conjunction with only one other country. Even so, neither France nor Spain helps, and each rebuts Italy's fairness complaints by saying that it is not imposing burdens on Italy because the other is also unwilling to contribute. This, of course, is just the *partial help* case. As I argued above, because both France and Spain fail to show proper regard for what is morally important

⁷⁹ The problem I identify here as the particularity problem has parallels in the literature on political obligation. There, the problem applies to theories grounded in the natural duty of justice. Political obligation seems to be owed primarily or exclusively to the institutions that apply to me, while the natural duty of justice seems to allow discretion regarding which institutions I support. For a defense of natural duty theories see Waldron (1993). For a statement of the particularity problem see Simmons (1979) chapter VI, and for a reply to Waldron see Simmons (2005) pages 170-179.

(the migrants' lives), and since these failures are sufficient to impose extra burdens on Italy, each plays a role in dumping the burdens Italy picks up. On these grounds, Italy has a fairness complaint against each country.

At this point, the discretionary nature of duties to rescue leaves open a further possible response for France and Spain. Suppose everyone knows that if Spain contributes to any rescue effort, it will be to one run by Bulgaria, not Italy. Thus, even if both France and Spain contribute their fair share to rescue efforts, Italy's burdens will not be lightened. In this way, France argues that it does not treat Italy unfairly because France's and Spain's failures to respond appropriately to the relevant reasons do not lead them to withhold contributions that would be sufficient to reduce Italy's burdens.⁸⁰ Since Spain has the discretionary freedom to choose which rescue effort to contribute to, neither it nor France counts as contributing to Italy's burdens. If this is the end of the story, there are simply more burdens to go around than can be borne. Full compliance with the duty to rescue would require maximum sacrifice from everyone required to make any sacrifice. In that case, France's argument goes through, and neither it nor Spain dumps any burdens on Italy. However, assuming that full compliance will not require maximal sacrifices from everyone, Italy's argument can take a further step to match the step taken by France's argument.

So far, I have presented the case as though Spain and France are the only (relevant) actors not doing their part. But this is an artefact of thinking of duties as non-discretionary. Now that we are thinking of a discretionary duty to rescue, the relevant pool of burdens is all the burdens associated with all the rescues that need to be performed and that require collective action.⁸¹

⁸⁰ See Feinberg (1984, 60-64) for a discussion of similar cases in the context of imperfect duties (duties that lack a prescribed time or place of fulfillment; these are precisely discretionary duties in my sense).

⁸¹ See Elinor Ostrom (2010) for a detailed analysis of collective action problems and the contexts in which they often arise.

Given this, the pool of potential contributors includes every agent—natural or artificial—who is bearing less than her fair share of the overall burdens. If we now imagine that no one fails to do her part through a blameworthy failure to be reasons responsive, we will imagine a scenario in which all these agents bear their fair share of the total pool of rescue-related burdens. If we have good reason to think that Italy's burdens would be reduced in *this* situation, then Italy has a fairness complaint not only against France and Spain, but against everyone who is not contributing her fair share to rescue efforts around the world. This is the *partial help* argument writ large.

So, even though it seems right to say that the duty to rescue allows for significant discretion on the part of those bound by it, those already rescuing almost certainly have legitimate fairness complaints against most non-contributors. Because they fail to be appropriately reasons-responsive, they play a role in (make a vector-contribution to) the unfair imposition of burdens by failing to bear their portion of the total pool of rescue-related burdens.

4. Omission and Collective Action

One might worry that the *partial help* argument has been writ too large. Consider the following case. Italy is rescuing migrants crossing the Mediterranean, Bulgaria is rescuing migrants crossing the land border from Turkey, and both efforts are on a par in all relevant respects. France, however, is rescuing no one. Italy knows that if everyone were doing her fair share of rescues, its burdens would be lighter than they currently are. Unfortunately, Italy can only influence France. Thus, Italy begins making fairness complaints against France, and France, exercising its discretion, begins contributing to Bulgaria's scheme. Nothing has changed for Italy, but since France is now doing its part, Italy no longer has a fairness complaint against it.

This seems odd. Italy claims to be treated unfairly by France because France plays a role in Italy's excessively heavy burdens. Yet France successfully satisfies its fairness obligation to Italy without reducing Italy's burdens. One might take this to show that the *partial help* argument is not ultimately concerned with fairness. If it were, it would argue that Italy's claim against France removes France's discretion so that it must contribute to Italy's rescue effort.

This line of objection can be interpreted as asserting one of two underlying thoughts. First, the objection might be another way of claiming that X treats Y unfairly by failing to contribute just in case X's contribution alone would be sufficient to reduce Y's burdens. My main argument up to this point has been aimed at rejecting this intuition, so I will set this interpretation aside. Alternatively, one could take the objection as expressing something like the following: if X treats Y unfairly by not contributing to any rescue effort, then it is also the case that X treats Y unfairly by contributing to any rescue effort other than Y's. So, in the EU example, since France could lighten Italy's burdens, France treats Italy unfairly when France contributes to Bulgaria's rescue scheme instead of Italy's.

To see where this second suggestion leads, suppose for the sake of argument that France imposes burdens on Italy when it performs no rescues *and* when it performs its fair share of rescues in Bulgaria's rescue effort. Granting this, it might seem to follow automatically that France treats Italy unfairly by contributing to Bulgaria's scheme.⁸² But how can this be? Recall that France's duty to rescue is supposed to be discretionary, and it seems clearly right to say that before Italy makes its complaint against France, France is free to contribute to either scheme. So, what changes when Italy makes its claim? Sarah McGrath gives us a potential answer in her

⁸² If this were right (and the ought-implies-can principle were true), it would be a serious problem for my view. If, for instance, Italy and Bulgaria announced fairness complaints against France at the same time on the same day, France would be forced to treat one of them unfairly (assuming it can only feasibly contribute to one scheme).

theory of causation by omission. She argues that omission “o causes [event] e iff o occurs, e occurs, and [commission of the act of which o is an omission] C_o is a normal would-be preventer of e” (McGrath 2005, 142).⁸³ A would-be preventer of e is something that would prevent e if it occurred. A would-be preventer is normal if it is *supposed* it to prevent e according to some *actual standard* (Ibid. 138). The thought is that Italy’s act of making an unfairness claim against France establishes a standard according to which France is supposed to help Italy and that this standard dissolves France’s discretion about which rescue efforts it may contribute to.

This proposal fails for several reasons. For one, it is not enough to simply establish a standard; the standard that is established must be shown to be important. McGrath’s notion of a standard is “of very general application”, covering “chess moves, dance steps, quiz answers, beliefs, baseball pitches, ways of beating eggs and stitching hemlines” (Ibid. 139). Each involves a standard of correctness that can be used to judge good and bad chess moves, dance steps, and so on. In that sense, all the standards are normative. However, they do not all have moral force. In fact, morality can be conceived as another standard by which actions can be judged to be appropriate or inappropriate. Since, according to the duty to rescue, France has moral discretion to contribute to whatever rescue effort it chooses, the standard established by Italy’s complaint will be ineffectual unless it can be shown to have overriding moral significance. Since the mere statement of the complaint does nothing to change the facts of the situation, it is unclear where this significance could come from.

Even if this difficulty could be overcome, problems still arise. Suppose France can only contribute to one scheme and Italy and Bulgaria make simultaneous fairness complaints against it, each demanding that France contribute to their rescue effort. To whose scheme should it

⁸³ McGrath offers a more precise formulation of the same principle, but this will do for my purposes here.

contribute? The most natural response to the situation is that France is free to choose which scheme to contribute to. In this case, its discretion persists. The only apparent alternative is to say that, even when France entirely fulfills its duty to rescue, by helping Italy for instance, it is still guilty of unfairly dumping burdens on Bulgaria. Surely this is a principle that *should* be rejected for imposing unfair burdens, though here the unfair burdens are placed on burden-dumpers rather than slack-takers.⁸⁴

Additionally, it is not clear why Italy's articulation of its fairness complaint should *create* a standard for France. Italy's speech act appears descriptive, not performative. It reports reasons to which France ought to respond; it doesn't create them. Thus, the standard according to which France treats Italy unfairly unless it contributes to Italy's rescue efforts applies whether Italy makes a declaration or not. But then, since Bulgaria is in the same position as Italy relative to France, it too must have an identical claim to France's contribution. So, France will have just as much reason to contribute to Bulgaria's scheme as to Italy's whether Bulgaria or France or anyone else makes a fairness complaint against it or not. France once again finds itself unfairly bound to shoulder more burdens than it can bear.

The initial objection was that something has gone wrong with the *partial help* argument since Italy's unfairness complaint against France, grounded in its unfairly heavy burdens, does not obligate France to contribute to Italy's rescue effort. Intuitively, we might think that if France treats Italy unfairly, it ought to contribute to Italy's scheme. But this intuition is misguided because its focus is too narrow. France isn't the only non-contributor, and Italy's scheme isn't

⁸⁴ One might attempt to run a similar omissions argument by appealing to a Lewisian view on which *o* causes *e* iff *C_o* would have prevented *e* (see Lewis 2000). By this standard, every agent in the world whose contribution to Italy's scheme would reduce its burdens, if it so contributed, counts as causing Italy's burdens by omission. But if this is right, we clearly have not landed on a normatively significant sense of omission. Suppose Bulgaria begins its scheme before Italy. The fact that Bulgaria causes by omission Italy's excessive burdens clearly does not mean that Bulgaria treats Italy unfairly or that Bulgaria ought to terminate its own scheme to contribute to Italy's.

the only one around. Still, one might try to vindicate the intuition by arguing that once Italy makes its claim on France, France counts as causing Italy's extra burdens by omission. As we have seen, however, this argument does not look promising.

5. Further Applications

In this chapter, I have argued that when we fail to contribute our fair share in a rescue effort and others must take up the slack, we treat those others unfairly. Problems we have a duty to solve that require collective action to address are subject to distributive norms that generate fairness obligations between rescuers in addition to the natural-duty obligations owed to those in need of rescue. The central objection to which I respond argues that one only dumps burdens when one's contribution alone would be sufficient to lighten the burdens of current contributors. I argue that this claim is mistaken. By failing to do one's part in the absence of excusing conditions, one fails to be appropriately reasons responsive. This failure makes one a blameworthy member of the vector group whose actions or omissions push in the direction of burden-dumping. Thus, those who fail to do their part are implicated in the resulting unfair distribution of burdens.

This argument appears problematic in the context of discretionary duties. When I am not obligated to contribute to any particular rescue effort, it is not clear who is treated unfairly when I do less than my fair share. I argue that this worry can be dispelled by broadening the scope of the argument. The argument I develop in response to the *partial help* case shows that one can play a role in the unfair burdens borne by individuals performing rescues even if one's contributions alone would not reduce their burdens. Thus, no matter how much discretion I have

in fulfilling my duty to rescue, I unfairly dump burdens on those who do their part when I fail to do mine.

It is worth considering how the arguments presented here apply in other contexts. Very briefly, I discuss voting and climate change. In the context of burden-dumping, voting and rescuing are fundamentally different because one cannot dump one's duty to vote on others. I shouldn't vote twice in an election because someone else didn't vote at all. This does not mean, however, that the arguments I have developed are inapplicable.

In its most general terms, the view I defend identifies responsible causal contributors to the outcomes of collective actions or omissions. This was the payoff of combining Goldman's vector theory of causation with Brink's and Nelkin's theory of moral responsibility. For any case in which we can identify the reasons to which individuals ought to respond, we can, in principle, identify those who are blameworthy (or praiseworthy) for the outcomes of their actions or omissions. In the case of voting, bad outcomes of elections or referendums can be very destructive even though no burden-dumping is involved. The account I have defended allows us to identify those who are blameworthy for pushing toward these negative outcomes even when, for instance, the better candidate wins. Burden-dumping can therefore be seen as a special case focusing on situations in which partial compliance affects the distribution of burdens. Many collective action problems are plagued by partial compliance and, in these cases, it is worth understanding how to assign blame and responsibility for unfair distributions.

Climate change is structurally much closer than voting to rescue cases and so raises similar distributive questions. Our responses to climate change, whether in the form of mitigation (preventing future climate change), adaptation (responding to unavoidable change), or compensation (to those unjustly affected), require collective action and allow burden-dumping.

Climate change raises additional complex questions about the initial fair distribution of burdens, intergenerational justice, cosmopolitanism versus nationalism, and so on.⁸⁵ However we answer these questions, the analysis presented here can help us respond appropriately to actors who fail to do their part.

Some duties are quite stringent, and this stringency can obscure distributive concerns. In rescue cases, for example, complaining that I must do more than my fair share of rescues when those being rescued are in dire need might seem melodramatic. It is worth remembering, however, that burden-dumping can impose very heavy burdens, especially when the duties involved are stringent. Additionally, those on whom the burdens fall may be better positioned to hold accountable those who refuse to do their part. This last point is especially relevant in the case of climate change.

There is widespread agreement that individuals acting independently cannot respond adequately to climate change.⁸⁶ Individuals, corporations, governments, and supranational organizations must act in concert if we are to minimize the damage of climate change to human well-being. But many actors are, and have been, unwilling to do their part; the US government, for instance, has consistently failed to pursue meaningful emissions-reduction policies.⁸⁷ While people outside the US often feel the effects of these failures most strongly, it is US citizens that can act most effectively change the trajectory of US policy. One strategy for pressuring the government is to voice complaints that the failure of the government (and others) to adequately respond to climate change has imposed unfair burdens on individual members of the population,

⁸⁵ For discussion of who should pay for the costs of mitigation, adaptation, and compensation, see Caney (2005). For discussions of intergenerational ethics related to climate change see Gardiner (2006) and Gosseries (2004).

⁸⁶ This is separate from the question of whether individuals have a duty to reduce their own emissions when others fail to act. For discussion, see Sinnott-Armstrong (2005), Schwenkenbecher (2014), and Hourdequin (2010).

⁸⁷ See Jamieson (2014) for an overview of the history of climate change.

requiring them to unilaterally reduce their private emissions or attempt to organize their own emissions-reduction schemes. Even the need to voice complaints is an avoidable and unfair burden. If we focus exclusively on the harmful effects of climate change, these grievances will go unnoticed. Not only does this let blameworthy actors partially off the hook, it robs those seeking change of a potentially important means of pressuring those who neglect their duties.

I have here only scratched the surface of the various ways in which the ethics of slack-taking and burden-dumping might be applied. My hope is that this discussion will help promote further applications by illuminating not just the structure of our duties to rescue, but a more general relationship between natural duties and fairness obligations.

Chapter 4, in full, has been submitted for publication of the material as it may appear in the *Journal of Ethics and Social Philosophy*, 2021, Finley, Aaron, Elsevier, 2021. The dissertation author was the primary author of this paper.

References

- Agnafors, Marcus. 2017. "On Disjunctive Rights." *The Southern Journal of Philosophy*, 55 (2): 141-157.
- Baron, Marcia. 1987. "Kantian Ethics and Supererogation." *The Journal of Philosophy*, 84 (5): 237-262.
- Björnsson, Gunnar. 2020. "Collective Responsibility and Collective Obligations Without Collective Moral Agents." In *The Routledge Handbook of Collective Responsibility*, S. Bazargan-Forward and D. Tollefsen (Eds.), 127-141. New York: Routledge.
- Brink, David O. 2013. "Situationism, Responsibility, and Fair Opportunity." *Social Philosophy and Policy*: 121-149.
- Brink, David O., and Dana K. Nelkin. 2013. "Fairness and the Architecture of Responsibility." In *Oxford Studies in Agency and Responsibility Vol. 1*, D. Shoemaker (Ed.), 284-313. Oxford University Press.
- Caney, Simon. 2005. "Cosmopolitan Justice, Responsibility, and Global Climate Change." *Leiden Journal of International Law*, 18: 747-775.

- Cohen, L. Jonathan. 1981. "Who is Starving Whom?" *Theoria*, 47 (2): 65-81.
- Feinberg, Joel. 1984. "The Moral and Legal Responsibility of the Bad Samaritan." *Criminal Justice Ethics*, 3 (1): 56-69.
- Gardiner, Stephen. 2006. "A Perfect Moral Storm: Climate Change, Intergenerational Ethics and the Problem of Moral Corruption." *Environmental Values*, 15 (3): 397-413.
- Goldman, Alvin. 1999. "Why Citizens Should Vote: A Causal Responsibility Approach." *Social Philosophy and Policy*, 16 (2): 201-217.
- Gosseries, Axel. 2004. "Historical Emissions and Free-riding." *Ethical Perspectives*, 11 (1): 36-60.
- Hope, Simon. 2014. "Kantian Imperfect Duties and Debates over Human Rights." *Journal of Political Philosophy*, 22 (4): 396-415.
- Horton, Keith. 2011. "Fairness and Fair Shares." *Utilitas*, 23 (1): 88-93.
- Horton, Keith. 2004. "International Aid: The Fair Shares Factor." *Social Theory and Practice*, 30 (2): 161-174.
- Hourdequin, Marion. 2010. "Climate, Collective Action and Endividual Ethical Obligations." *Environmental Values*, 19 (4): 443-464.
- Igneski, Violetta. 2006. "Perfect and Imperfect Duties to Aid." *Social Theory and Practice*, 32 (3): 439-466.
- Jamieson, Dale. 2014. *Reason in a Dark Time: Why the Struggle Against Climate Change Failed—And What it Means for Our Future*. Oxford: Oxford University Press.
- Kamm, Frances. 1999. "Famine Ethics: the Problem of Distance in Morality and Singer's Ethical Theory." In *Singer and his Critics*, D. Jamieson (Ed.), 162-208. Oxford: Blackwell Publishers Ltd.
- Karnein, Anja. 2014. "Putting Fairness in its Place: Why There is a Duty to Take up the Slack." *The Journal of Philosophy*, 111 (11): 593-607.
- Lewis, David. 2000. "Causation as Influence." *Journal of Philosophy*, 97 (4): 182-197.
- Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly*, 2 (4): 245-264.
- McGrath, Sarah. 2005. "Causation by Omission: A Dilemma." *Philosophical Studies*, 123 (1-2): 125-148.

- Miller, David. 2013. "Taking Up the Slack? Responsibility and Justice in Situations of Partial Compliance." In *Justice for Earthlings*, D. Miller (Auth.), 206-227. Cambridge: Cambridge University Press.
- Murphy, Liam. 2000. *Moral Demands in Nonideal Theory*. New York: Oxford University Press.
- Murphy, Liam. 1993. "The Demands of Beneficence." *Philosophy and Public Affairs*, 22 (4): 267-292.
- Ostrom, Elinor. 2010. "Beyond Markets and States: Polycentric Governance of Complex Economic Systems." *American Economic Review*, 100 (3): 641-72.
- Pogge, Thomas. 2011. "Are We Violating the Human Rights of the World's Poor?" *Yale Human Rights and Development Journal*, 14 (2): 1-33.
- Ridge, Michael. 2010. "Fairness and Non-Compliance." In *Partiality and Impartiality: Morality, Special Relationships, and the Wider World*, B. Feltham and J. Cottingham (Eds.), 194-222. Oxford: Oxford University Press.
- Schwenkenbecher, Anne. 2014. "Is There an Obligation to Reduce One's Individual Carbon Footprint?" *Critical Review of International Social and Political Philosophy*, 17 (2): 168-188.
- Sher, George. 2009. *Who Knew?: Responsibility Without Awareness*. New York: Oxford University Press.
- Wellman, Christopher, and A. John Simmons. 2005. *Is There a Duty to Obey the Law?*, 121-188. Cambridge: Cambridge University Press.
- Simmons, A. John. 1979. "The Natural Duty of Justice." In *Moral Principles and Political Obligations*, A.J. Simmons (Auth.), 143-156. Princeton: Princeton University Press.
- Singer, Peter. 2007. "Famine, Affluence and Morality." In *Ethical Theory: An Anthology*, R. Shafer-Landau (Ed.), 505-512. Blackwell.
- Sinnott-Armstrong, Walter. 2005. "It's Not My Fault: Global Warming and Individual Moral Obligations." In *Perspectives on Climate Change: Science, Economics, Politics, Ethics*, W. Sinnott-Armstrong and R. Howarth (Eds.), 285-307. Bingley: Emerald Group Publishing Limited.
- Smith, Angela. 2008. "Control, Responsibility, and Moral Assessment." *Philosophical Studies*, 138 (3): 367-392.
- Smith, Patricia. 1998. *Liberalism and Affirmative Obligation*. Oxford: Oxford University Press.
- Talbert, Matthew. 2016. *Moral Responsibility: An Introduction*. Malden: Polity Press.

Unger, Peter. 1996. *Living High and Letting Die*. New York: Oxford University Press.

Waldron, Jeremy. 1993. "Special Ties and Natural Duties." *Philosophy and Public Affairs*, 22 (1): 3-30.

Wolterstorff, Nicholas. 2008. *Justice: Rights and Wrongs*. Princeton: Princeton University Press.

Chapter 5

Multiple Principles of Political Obligation

Introduction

In chapters 1 and 3 I addressed the acceptance and cooperation objections to the principle of fairness (PoF), but we might still wonder when the benefits of a scheme are worth their cost. This question has a standard answer: benefits are worth their costs when the recipient prefers receiving the benefits together with the costs to receiving neither. However, this answer does not tell us how many people do receive net benefits from government activities. It also fails to specify which activities are beneficial, and to whom. This uncertainty leaves open the possibility that relatively few members of political communities receive benefits that are worth their cost, and therefore that relatively few have political obligations under the PoF.

If the PoF binds only a few citizens to obey the law, it violates an adequacy condition for theories of political obligation known as *generality*. This condition requires that all, or nearly all, citizens be bound to obey all, or nearly all, the laws of the country in which they reside.⁸⁸ The core of the generality problem facing the PoF comes from its requirement that benefits must outweigh burdens and that the most common way to measure benefits and burdens is in terms of preference satisfaction. Actual preferences are highly diverse, are sometimes unusual, and may be irrational (due to ignorance or misinformation, cognitive biases, adaptive preferences, weakness of will, and so on). If the configuration of preferences within the citizenry fails to align well with the public goods produced by the government, those goods may not be worth their cost for a substantial portion of the population.

⁸⁸ I set aside here the potential complication of alien resident citizens.

One response to this situation would be to appeal to idealized preferences—preferences “I would have if I were to engage in thoroughgoing deliberation about my preferences with full pertinent information, in a calm mood, while thinking clearly and making no reasoning errors” (Arneson 1989, 83). These may be more plausible than actual preferences as an account of individual well-being, but they do not help here unless we assume that the process of idealization leads preference to converge in ways that the objection suggests actual preferences do not.⁸⁹ To satisfy the generality condition, idealized preferences must converge such that the goods and services provided by government are worth their cost for most individuals. However, it is not obvious that the process of idealization would lead preferences to converge rather than diverge further. Perhaps if I were to engage in the kind of thoroughgoing deliberation described above, I would discover that most of my attitudes toward the government derive from a kind of false consciousness that suppresses my true preferences. In any case, the appeal to idealized preferences exchanges one set of preferences for another, but without any obvious impetus toward convergence.

Another response would be to focus on preferences that everyone can be presumed to have. This is the route George Klosko takes in focusing on what he calls ‘presumptive goods’, along with whatever is necessary to provide them (Klosko 2004). Presumptive goods, like Rawls’s primary goods, are goods everyone can be presumed to want whatever else they want. This, however, severely limits the scope of the PoF, restricting it to goods such as national defense, public security, or the rule of law, along with the infrastructure, such as roads and manufacturing, necessary to provide those benefits. These are central governmental functions to be sure, but, Klosko claims, they do not cover a wide range of government activities including

⁸⁹ For further discussion of preference-based theories of well-being, see (Arneson 2016; Bykvist 2016).

“services that benefit *other* people [such as]...the poor, handicapped, or otherwise disadvantaged,” regulating the economy, supporting recreation and culture through public parks, funding for the arts, and so on, all of which, “although highly desirable, [are] arguably neither indispensable for satisfactory lives nor necessary for the provision of essential public goods” (Klosko 2019, 72-73 original emphasis).⁹⁰

Thus, the dependence of the PoF on the balance of benefits and burdens as measured in terms of preference satisfaction requires us either to depend on empirical data (or idealizations of empirical data), in which case it is unclear that the generality requirement for political obligation will be satisfied, or it requires us to so severely limit the scope of the PoF that everyone is bound to obey only a few laws.

A principle that requires citizens to obey only a few laws cannot support an adequate theory of political obligation. Instead, a theory must meet the *comprehensiveness* condition, which requires “that a theory must establish moral requirements to obey all [or nearly all] laws passed by the state” (Klosko 2019, 65). Standardly, the PoF applies only to schemes that produce public goods. While Klosko dramatically underestimates the extent to which society is permeated by public goods and schemes to provide them,⁹¹ it is not clear that all intuitively

⁹⁰ Klosko’s focus on presumptive goods is almost certainly too narrow. He claims, for instance, that welfare programs “that support the poor, handicapped, or otherwise disadvantaged” can only be seen as public goods “to the extent that they keep the poor minimally satisfied and so not disruptive of public order” (Klosko 2019, 72-73). However, having a larger population of content and productive people will contribute to the welfare of society as a whole. They may be productive in the economy, generate culture, and so on. Furthermore, just as the availability of public health care is beneficial to those who aren’t sick because it provides security in case they need it, welfare programs provide security against unexpected misfortunes that would otherwise plunge one into poverty. Similar points could be made about managing the economy or supporting the arts.

Simmons presents another kind of objection, arguing that, at bottom, Klosko’s theory has the character of a natural duty theory rather than a fairness theory. The focus on presumptive goods shifts concern from unfair advantage-taking to the intrinsic importance of these goods and of seeing them provided. This concern, as Simmons notes, seems more at home in a theory based on natural duties (Simmons 2001, 35).

⁹¹ cf. Arneson 2013 who argues that the economy and much of society is “thickly marbled by obligations stemming from the principle of fairness” (139).

legitimate government activity can be justified in this way. So, the PoF faces two challenges, one related to generality, the other related to comprehensiveness.

Klosko's response to both challenges is to turn to what he calls *Multiple Principle* (MP) theory.⁹² As the name suggests, MP theories combine multiple principles of political obligation, and Klosko combines the PoF with a duty of mutual aid—similar to Christopher Wellman's strategy which will be explored in section 3.2. In this chapter I construct my own MP theory by combining the PoF with natural duties, and I follow Rawls and Waldron in making use of the natural duty of justice rather than a duty of beneficence or rescue.

The natural duty of justice requires us to not undermine just institutions anywhere, and to support just institutions or help establish them where none exist, so long as this is not too costly.⁹³ This approach to political obligation does not face generality problems since natural duties, by definition, bind everyone. Instead, it faces the *specificity* problem. A successful principle of political obligation must not only generate duties to obey the law, it must generate duties to obey the law *of one's own country* (or the country one resides in). The natural duty of justice struggles to do this. Because it generates a duty to support or establish just institutions as such, it is not clear how it could require us support or establish only those institutions that “apply to us”, to use Rawls's phrase (Rawls 1999, 99).

So, while the PoF threatens to apply too narrowly in terms of the people in whom it generates obligations and in terms of laws to which it applies, the natural duty of justice threatens to apply too broadly. I will argue that, by uniting the two, this cluster of problems can

⁹² Another response would be to adopt philosophical anarchism and maintain that, even if governments' use of coercive force is justified, no individual (or very few) has an obligation to obey the law. I argue that this conclusion should be rejected.

⁹³ I set aside the important question of how costly is too costly. Here, I only assume that the demands of justice are small enough to give us significant room to pursue the lives we choose. I also set aside questions of individual justice to focus on institutional justice. While individual justice is vitally important, a theory of political obligation must focus on the justice of institutions.

be avoided. The natural duty of justice will serve as the fundamental principle, and the requirement that we fairly distribute the benefits and burdens of instituting and maintaining just institutions will allow us to invoke the PoF. The PoF will solve the specificity problem by binding the beneficiaries of just institutions to their fellow citizens, and the natural duty of justice will solve the generality problem by imposing political obligations independently of benefits received. Similarly, the natural duty of justice (combined with a fully developed theory of justice) tells us which institutions we have a duty to support. These institutions might produce public benefits, or they might not. Thus, the natural duty of justice effectively decouples the PoF from its publicity requirement. In section 1, I discuss the way in which multiple principles need to be combined in order to be mutually reinforcing. In section 2, I address the specificity problem by combining the natural duty of justice with the PoF. I also consider Jeremy Waldron's and Christopher Wellman's prominent attempts to solve the particularity problem. In sections 3 and 4 I argue that my combination of principles meets the generality and completeness requirements, and in section 5 I briefly discuss the general notion of reciprocity and a lingering worry about specificity.

1. Multiple-Principle (MP) Theory

Jonathan Wolff and George Klosko have both argued that political obligations can be best grounded in a theory that appeals to multiple principles. There are a number of ways a theory can combine multiple principles, and the most straightforward is simply to affirm more than one as true. This strategy does not lead to interaction effects between principles but can use different principles to do different jobs within the theory. In section 3.2, for instance, I argue that Wellman uses the natural duty of justice to support the right of governments to rule, and the PoF

to support the obligation of citizens to obey the law. I will call this a *patchwork* approach to MP theory (cf. Wolff 1995, 14-15).

The patchwork approach is most useful for answering objections to the ability of a particular principle to solve independent problems of political obligation and legitimacy. Wolff, for instance, argues that different principles are needed to legitimate different functions of government (Wolff 1995, 18). Or, one might argue that different portions of a population are bound to the same set of laws by different principles (or both). Either way, MP theory might be able to answer objections arguing that individual principles cannot generate genuine political obligations because they either do not bind enough of the population, or they do not bind the population to enough of the laws and institutions that purport to govern them. More generally, if questions about political obligation and legitimacy break down into a patchwork of independent problems, a patchwork of independent principles might provide an appropriate way to address them.

While I do not reject patchwork theories as such, I also think they will not provide the main substance of a theory of political obligation. Theories must face objections that the patchwork approach is not well equipped to address. To illustrate, standard objections to fairness and consent theories argue that if they produce political obligations for anyone, they generate them for a small minority of the population. If so, combining these principles won't satisfy a generality condition that requires a substantial majority of the population to have political obligations. The natural duty of justice (and other natural duties), by contrast, apply too broadly. If I have a duty to promote justice, it is unclear why I must promote it here rather than in some distant land. Adding independent obligations to the theory will not solve this problem. Gratitude theories, unlike consent, fairness, and natural duty theories, appear to be non-starters. As

Simmons notes, “it is clear, first, that debts of gratitude are only owed to benefactors who have acted with certain motives and, second, that our benefactors are not entitled to themselves specify what shall count as a fitting return” (C. Wellman & Simmons 2005, 119).⁹⁴

These objections are not ones that can be solved simply by affirming several principles at once. On the face of it, a successful patchwork theory of political obligation must appeal to new principles, or must somehow reframe the debate so that political obligation and legitimacy break down into a set of new questions that the familiar principles *can* answer. Still, since a patchwork theory could in principle lead to a viable theory of political obligation, it will be instructive to briefly consider one attempt to defend a patchwork theory.

In his 1995 paper on MP theory, Jonathan Wolff briefly outlines what he takes to be a promising patchwork theory of political legitimacy and obligation. First, he breaks government activity into four categories: (i) providing security, both internally and from foreign invasion; (ii) providing public goods that are beneficial to all; (iii) providing public goods that are beneficial only to some; (iv) redistributing wealth and income (Wolff 1995, 18). Wolff argues that, in each category, whatever principle legitimates government authority also generates obligations to obey the law.

For category (i), Wolff appeals to a principle of self-interest, though he acknowledges that the PoF (which he applies to category (ii)) might also cover this category. Considerations of self-interest, however, support at most prudential obligations to obey the law, not moral ones. They might legitimate governmental authority, but this leaves us with philosophical anarchism. The government may enforce laws, but we have no obligation to comply except where independent moral and prudential considerations apply. Similarly, Wolff claims that category

⁹⁴ For further discussion see Simmons 1979, chapter VII.

(iii), if it can be supported at all, should be subsumed into either category (ii) or (iv), each of which gets its own principle—the PoF and the natural duty of justice respectively. Wolff’s patchwork theory, then, affirms two principles each of which covers a separate function of government.

Wolff’s discussions of the PoF and the natural duty of justice are very brief. His aim is only to propose a brief sketch of a theory of political obligation in a paper mainly focused on exploring the conceptual space of MP theory. For (ii), he assumes a non-voluntarist interpretation of the PoF and states that the burden of proof lies on recipients to show that they do not benefit from the receipt of “widely accepted goods” (Wolff 1995, 23). For (iv), Wolff appeals to the natural duty of justice as the basis for a theory of distributive justice. He briefly explains how such a theory might legitimate political authority, but does not mention the particularity problem. So here, as in (i), we appear to be left with, at most, a defense of philosophical anarchism.

What is clear is that limiting the application of each principle to only some functions of government does not strengthen the overall defense of political obligations. The arguments against each principle remain effectively untouched. The PoF, it is argued, generates obligations for very few *within the domain to which it applies*. For someone like Simmons, the problem with the PoF is not that it is applied to the wrong domain, but that it generates too few obligations within its domain. Similarly, the problem with natural-duty theories is not that they cannot generate obligations to support just public-goods schemes, but rather that they cannot generate obligations to support *particular* public-goods schemes.

Given the lack of interaction between principles, an argumentative strategy common among philosophical anarchists such as Simmons and Michael Huemer is left unaddressed.

Klosko calls this the “divide and conquer” strategy (Klosko 2019, 68)), and it proceeds by considering each potential principle of political obligation separately.⁹⁵ After arguing against one principle, it is set aside as inadequate and the next principle is taken up. Patchwork theories such as the one outlined above cannot respond to this strategy. Simmons states this explicitly.

The failures of... [standard theories of political obligation] to account independently for the duty to obey [the law] infect any efforts to combine the insights of the various families [of theories] ... Only if each family could account for the duties of a distinct significant minority of citizens could they hope to account collectively for a general duty to obey. My criticisms... are not, however, consistent with such a supposition. (Wellman & Simmons 2005, 102)

Simmons clearly has patchwork theories in mind, and I am inclined to agree with his criticism. Thus, a more adequate MP theory must combine principles in ways that modify the internal structure of the principles themselves.

2. The Natural Duty of Justice

The theory I develop here is grounded in the natural duty of justice. This duty requires us not to undermine just institutions, to support just institutions where they exist, and to help establish just institutions where they do not exist. This version of the duty is notably different from those defended by Rawls, Waldron, and Kant. Rawls and Waldron argue that the natural duty of justice requires us to support institutions that apply to us (either our duty is owed exclusively to local institutions, or we must prioritize the support of these institutions),⁹⁶ and Waldron and Kant

⁹⁵ Prominent examples of this argumentative strategy can be found in Simmons (1979) and Huemer (2013).

⁹⁶ As Rawls states the duty, it “requires us to support and comply with just institutions that exist and apply to us. It also constrains us to further just arrangements not yet established, at least when this can be done without too much cost to ourselves” (Rawls 1999, 99). His inclusion of the phrase “apply to us” effectively circumvents the particularity problem as described below, but the condition cannot be taken for granted and does not fit well with the general structure of natural duties. Waldron’s defense of natural-duty theories of political obligation (discussed below), for instance, is mainly an elaboration and defense a version of Rawls’s application clause.

argue that we are only obligated to establish just institutions with those close to us to whom we pose an immanent threat.

Without these restrictions, the natural duty of justice falls prey to the *particularity problem*. Natural duties have two defining features. For one, they apply to us simply because we are moral agents, not because of our voluntary actions or special relationships (cf. Rawls 1999, 99). Additionally, they are owed to everyone. A promise places me in a moral relationship only with the person to whom I make the promise while a natural duty to rescue places me in a moral relationship with everyone. If they are in need and I am in a position to rescue them, I have a (defeasible) duty to do so. Thus, natural duties seem to lack the resources to systematically and reliably generate duties that are limited in the ways a theory of political obligation needs them to be.

Jeremy Waldron and Christopher Wellman have both presented influential defenses of natural-duty theories aimed at answering the particularity problem. As we will see, both arguments rely on importing additional particularizing considerations—and in Wellman’s case, I will argue, importing the PoF. I consider each in turn.

2.1 Waldron

Waldron begins with a compelling argument for the conclusion that there is a natural duty of justice, and that it is theoretically prior to other forms of political obligation. He observes that we have a duty not to undermine just institutions with which we have no contact. Residents of France, for instance, have a duty not to undermine just institutions in New Zealand.⁹⁷ This is true

⁹⁷ Waldron tells the story of the Rainbow Warrior Affair in which French agents sank a vessel named “Rainbow Warrior”, a Greenpeace vessel being used to disrupt French oceanic nuclear tests. French operatives bombed the vessel in Auckland Harbor, and, during the ensuing investigation by the New Zealand government, “officials of the

even for those who have never consented to the authority of just institutions in New Zealand, and who owe them no debts of reciprocity or gratitude. The only way to explain this duty, it seems, is to appeal to a natural duty of justice. And if this duty applies to just institutions in other countries, it also applies to just institutions in one's own country. Having established the existence of the natural duty of justice, Waldron's theory must go on to refute the objection that "such a duty cannot by itself account for the special [particular] character of political obligation" (Waldron 1993, 11).

The basis of Waldron's solution to the particularity problem is Kantian in inspiration. He adopts Kant's idea that, in the state of nature, we must quickly form a political society with those close to us. The state of nature is characterized by violence and the fear of violence since "individual men...can never be certain that they are secure against violence from one another" (Kant as quoted in Waldron 1993, 14). Because we are a special threat to those close to us, we have a special obligation to live with them according to common principles of justice. These principles require institutional administration and enforcement, which is best carried out by exactly one set of institutions per territory. In very broad strokes, the natural duty of justice requires us to establish local principles of justice that are administered by local institutions. Since we are duty-bound to live according to the principles, we are also duty-bound to submit to the rule of the institutions that enforce them.

To examine Waldron's theory in more detail, we must also examine the particularity problem in more detail. The problem has several dimensions, each of which Waldron's theory aims to address. The first is about the geographic reach of a principle. We can say a principle is *boundless* if it is geographically unlimited. By contrast, following Waldron, we can say a

French government living and working in France conspired to undermine the operation of the criminal justice system in New Zealand" (Waldron 1993, 9).

principle is *range-limited* if it only covers limited geographical areas. Waldron's theory is most explicitly concerned with defending the idea that principles derived from natural duties can be range-limited. To reach this conclusion, Waldron distinguishes between the natural duty of justice as such and particular principles of justice it supports, which brings us to the second dimension of the particularity problem.

This dimension is about who a principle applies to. We can say a principle is *automatic* if it presumptively applies to everyone in the geographical area it covers (unless someone can be shown to fall into an exception category). By contrast, we can say a principle is *conditional* if it presumptively does not apply to anyone unless they can be shown to satisfy a trigger condition (the PoF, for instance, does not trigger unless one receives benefits that are worth their cost). For Waldron, the trigger condition for range-limited principles of justice (or the trigger for the duty to decide on and institute such principles) is posing a threat to those close by. If someone does not pose a threat to anyone, she still has a duty not to undermine justice, and promote or create just institutions, but she will not be bound by any range-limited principles of justice. But even if she is bound, there may still be an open question about how she ought to do her part within the institutions that administer and enforce the principles she bound by.

This takes us to the third dimension of the particularity problem which is about how the requirements of a principle (or duty) are satisfied. We can say a principle is *discretionary* if it is up to the person under the obligation to satisfy it however she sees fit. We can say a principle is *non-discretionary* if the obligation requires the obligor to satisfy it in some specified way.

Defending a duty to obey the law requires defending a non-discretionary obligation. If Waldron's theory only concluded that we must support domestic just institutions, but left us the discretion to decide *how* to support those institutions, many people might have no particular reason to obey

the law. I might donate to advocacy groups rather than pay taxes, or create inspiring but illegal graffiti art in public places that improves public perception of the government. This is why Waldron argues that being bound by range-limited principles of justice implies submitting to the authority of the institutions that administer and enforce them. Without this argument, Waldron's theory would not be a full defense of the duty to obey the law.

To summarize, natural duties, including the natural duty of justice, are, at bottom, *BAD* (boundless, automatic, and discretionary). This is because they apply to all persons regardless of special relationships or institutional arrangements, and identify a broad end to promote without specifying the means by which it should be promoted. Waldron, however, argues that the natural duty of justice gives rise to *RCN* obligations (range-limited, conditional, and non-discretionary). Posing a threat to those close to us creates a conditional special relationship that can support range-limited principles that require non-discretionary submission to the authority of domestic institutions of justice.

The idea that we are a special threat to those close to us is the linchpin of Waldron's argument. This is the piece of the argument that establishes a special relationship between compatriots that removes (or at least limits) their discretion in fulfilling the natural duty of justice. As A. John Simmons has argued, however, this piece of the argument is unconvincing. Two criticisms go to the heart of the matter. The first has to do with actual versus felt security. In the quotations given above, Kant asserts that the state of nature will be characterized by both violence and the fear of violence. I know that there are some bad apples out there, so without institutions to maintain peace and security, I can't trust anyone I meet since anyone might be one of the bad ones. Simmons notes, however, that "we do not wrong others simply by virtue of their

subjective *feeling* of being threatened by us” (C. Wellman & Simmons 2005, 175 original emphasis).

Kant’s, and by extension Waldron’s, argument requires it to be true that I wrong you if I refuse to join institutions of justice with you. But to actually wrong you, I must actually pose a threat. In the state of nature it may be rational to live in fear of everyone, but that fear may not often be justified. Thus, the argument only requires those who are actually disposed to violence to enter into a political society with each other, leaving everyone else free to arrange their affairs as they see fit. This seems sensible from the perspective of limiting both violence and fear since, if we are confident that those disposed to violence are under the power of a government, anyone I meet not under such power is not likely to be violent. For this reason, Waldron’s argument does not convincingly meet the generality requirement.

Supposing that this objection could be met, another problem would face the Kant/Waldron account. The argument assumes that after entering a political society, one (and one’s children?) will remain under its authority so long as one remains within its jurisdiction. However, if the only objective of these institutions is to protect us from the violence posed by those close to us (including protection from foreign invasion), their authority, and our duty to submit to it, should last only as long as our threats to each other remain credible. In a modern society, however, I can be confident that the people I meet on a daily basis pose no immediate threat to me. These are people with whom I am most likely “to have established bonds of friendship and sociability”. Furthermore, “surely other states and their evil (e.g. terrorist) residents constitute greater threats to us than do harmless local nonparticipants [those who harmlessly disobey the law]” (C. Wellman & Simmons 2005, 175). If this is true, then we have no particular reason to remain subject to domestic institutions. Maybe we have a special duty to

prevent their collapse, but this should generate no more than weak and partial duties to obey the law. So long as one does not pose a threat to anyone else, one should have no obligation to obey the law. Once again, the Kant/Waldron account does not convincingly meet the generality requirement, and now fails to meet the comprehensiveness requirement as well. To the extent that a law does not serve to prevent violence, this argument cannot explain why we ought to obey it.

Waldron's general strategy is correct. To defend a natural-duty theory, we must find a special condition that converts the boundless and discretionary natural duty of justice into a range-limited and non-discretionary duty to obey the law. What we have seen is that posing a threat to those close to us is an inadequate condition.

2.2 Wellman

Christopher Heath Wellman has also developed a detailed natural-duty theory which he calls *Samaritanism*. Samaritan duties are duties to rescue, and, in the context of a duty to obey the law, Wellman focuses on duties to rescue others from the perils of the state of nature which "would be a horribly perilous environment" (C. Wellman & Simmons 2005, 11 cf. p.23). Since coercive political institutions present the only means of escape, and the benefits they provide are worth their costs, they have legitimate authority to coerce us, and we have a moral duty to obey them.

Wellman's theory seems initially very similar to Waldron's. The state of nature is dangerous primarily because we pose a threat to each other, and it is avoiding this threat that motivates the establishment of political institutions. But Wellman addresses the particularity problem differently than Waldron. Most notably, he appeals to independent moral considerations to argue that our Samaritan duties of rescue can be range-limited on the one hand, and non-

discretionary on the other. Wellman first argues that duties to rescue can be range-limited when the rescue constitutes a coordination problem. As he says, “political instability is fundamentally a coordination problem because there is no way to effectively eliminate the sources of this type of peril without coordinating people” (C. Wellman & Simmons 2005, 38). He contrasts this with a problem like famine, which, in principle, a very wealthy person (his example is Bill Gates) could solve unilaterally. However, Bill Gates cannot literally end famine unilaterally, no matter how wealthy he is. Rather, he uses his money and the institutions of states and the global economy to secure the coordination of others in combating famine. In this way, famine, just like political peril, cannot be eliminated without coordinating people.

What Wellman seems to have in mind are not coordination problems as such, but coordination problems that require local participation. Immediately after the Bill Gates example, Wellman asks why Oxfam America can’t coerce me into donating money to fight starvation. His answer is that “there is nothing special about *my* money—funds from Bill Gates or anyone else would be just as effective” (C. Wellman & Simmons 2005, 39 original emphasis). I can’t be coerced because my contribution is not necessary. What Oxfam needs is money, and money can come from anywhere. Political instability, however, is a local problem that must be solved within the imperiled community. Bill Gates might swoop in with his billions, but his attempts to solve the problem will be ineffective unless the people within the community go along with his plan to establish stable institutions. Wellman’s claim, then, is that when it comes to politics, domestic problems require domestic solutions.

For Wellman, Samaritan duties to rescue others from political peril is a range-limited duty in the sense that it applies community by community. Wellman frames this as a point about coercion. My government is justified in coercing me and my compatriots because our

cooperation is vital for solving problems of political instability, and that cooperation can only be reliably secured through coercion. Thus, range-limited principles are established at the institutional level, but not the individual level. The question for individuals is about discretion. Institutions of justice ought to establish and enforce range-limited principles of justice, but do individuals have the discretion to promote justice however they see fit even if that means disobeying the law?

The answer to this question has to do with *discretion*. Wellman argues that discretion is a good, but each of us is entitled only to a fair share of it. He imagines a scenario in which five generous neighbors all make enormous contributions to rescuing others (or, we can imagine, to promoting justice), but refuse to obey the law in various ways, say, by avoiding taxes. Wellman argues as follows.

[My neighbors] are able to disregard our state's legal commands only because the rest of us *do* respect the state's authority. More importantly, the rest of us would also prefer to ignore our government's legal commands in order to focus our attention more fully on the particular causes about which we feel most passionately, but the consequences would be catastrophic if all of us did so. In other words, my five neighbors are free riders *even though they are doing more than the rest of us* because they are enjoying the good of discretion, a good that is available only because the rest of us obey the law. (Wellman & Simmons 2005, 42 original emphasis)

Paradoxically, Wellman converts discretionary duties to rescue into non-discretionary obligations to obey the law by appealing to the value of discretion. We might wonder, however, why it is the sacrifices of these particular people that matters. After all, members of other political communities also sacrifice their discretion, so what binds me to my political community rather than theirs? The answer seems to be that it is my compatriots who make the good of discretion available to me in the first place. This is odd for two reasons. First, natural duties are generally understood to be discretionary, even in the state of nature. Discretion, here, might be

interpreted as a negative liberty. It is permissible for me to fulfill my duty in whatever way I want, of the options available to me, without restraint from others. Taken this way, no one can provide me with discretion. Others' actions can only restrict my discretion. Wellman, then, must have a more positive conception of discretion in mind. The thought might be that what matters is the number and quality of available options, and that in the state of nature my options are so few and so poor that I am effectively unable to exercise my discretion.

The positive interpretation seems to make sense of Wellman's claim that others' sacrifices provide the good of discretion. Then, once the good of discretion is available, giving it up is a sacrifice and, since others make this sacrifice, it would be unfair for me not to make it too. However, if I do make the sacrifice, I do not actually receive the good of discretion. So, in what sense am I made better off when my compatriots sacrifice their discretion? If anything, I seem to be worse off. In the state of nature, discretion simply wasn't available. Now that it is, I must face the frustration of forgoing it. Far from rewarding me with the carrot of discretion, political community and the sacrifices of my compatriots discipline me with the rod of abstinence from discretion. So, this interpretation is unable to explain why it is the sacrifices of my compatriots (as opposed to others' sacrifices) that generate my duty to obey the law. A broader interpretation of the good of discretion is more promising. Wellman notes as he introduces the idea that discretion is a good that "we would like to decide for ourselves which (if any) religion to practice, which (if any) profession to pursue, which (if any) people with whom to associate, and which (if any) hobbies to explore" (C. Wellman & Simmons 2005, 41). Even if I may not exercise discretion over how I fulfill my duties to rescue (or my duties of justice), I am able to exercise far more discretion over many other aspects of my life than I would in the state

of nature. Because I receive this significant benefit from the sacrifices of my compatriots, it would be unfair of me to take even more discretion on top of this.⁹⁸

Wellman's discussion of discretion employs the PoF, but limited to the good of discretion. His argument in the passage quoted above is grounded in a no-free-riding intuition, and he argues that others have a right to our compliance with a mutually beneficial scheme because they make sacrifices that benefit us all. We might further note that discretion, as Wellman understands it, is a public good. The ongoing maintenance of just institutions generates many meaningful options, and we benefit from the availability of these options. Wellman's defense of political obligation has all the hallmarks of the PoF. Why then present it as a natural-duty theory?

One answer is surely that Wellman elsewhere rejects fairness theories of political obligation (C. H. Wellman 2001, 737-38). If he were here endorsing a fairness-based theory, he would put himself in the awkward position of affirming blatantly inconsistent arguments.⁹⁹ Wellman could respond by rejecting his earlier arguments against the PoF, and given my own support for the principle, I think this is the best option available to him.¹⁰⁰ Thus, Wellman's theory of political obligation is best understood as an MP theory that employs natural duties and the PoF. But in Wellman's account, the two principles do not substantively interact. The natural

⁹⁸ Wellman's argument relies on the idea that discretion is always a good. To the extent that discretion can be read as freedom or liberty, this idea has been rejected, I think decisively, in Nussbaum (2003, 43-50) and Kymlicka (2002, 141-145). If Wellman's claim is that we have an interest not in discretion as such, but in discretion about important aspects of our lives, Wellman must argue that discretion over which political institutions to support is important.

⁹⁹ Simmons, in his critique of Wellman's argument, suggests that Wellman can avoid the PoF if his appeal to fairness derives from a duty to rescue according to which "all persons are naturally bound to their *local* political groups and the local tasks of those groups (C. Wellman & Simmons 2005, 186 original emphasis). As Simmons notes, however, this is clearly circular. Rather than explaining range-limited and non-discretionary duties, it assumes them.

¹⁰⁰ Interpreted this way, Arneson's 2013 response to Wellman's 2001 critique of the PoF can be taken as friendly support of Wellman's 2005 account of political obligation.

duty to rescue applies to governments and justifies their use of coercion while the PoF applies to members of political communities and explains their duty to obey the law independently of their Samaritan duties to rescue.

Wellman might respond by emphasizing that the type of discretion he focuses on is discretion in fulfilling one's natural Samaritan duties. This ties any fairness considerations to the natural-duty portion of the theory. The two do not operate independently. However, this link is not as tight as it seems. Wellman characterizes discretion as a *general* good, not as a good related specifically to duties to rescue. So, while he does emphasize giving up discretion in how one performs rescues, any sacrifice of discretion would do. So long as others in the community give up some important discretion or other, and I benefit, I will have an obligation to similarly sacrifice some of my important discretion. And while the existence of a political community provides a lot of opportunities that otherwise would be unavailable, *my* following the law within a political community limits my discretion over, as Wellman mentions, my religion, profession, associates, and hobbies. Even if obeying the law had nothing to do with Samaritan duties to rescue, limitations on these aspects of discretion would, for Wellman, trigger fairness obligations that I do my part.

Thus, Wellman's MP theory is patchwork theory that remains vulnerable to the divide-and-conquer strategy discussed earlier. Despite his attempts to introduce ideas of fairness, the natural duty of justice is left to operate in isolation.

2.3 Justice and Specificity

While Wellman's theory rightly raises questions of fairness about the distribution of benefits and burdens produced by fulfilling natural duties, it does so in a way that does not alter

the structure of the natural duty he considers or the fairness considerations he raises. This is the crucial theoretical innovation of the theory I present here.

Wellman argues that Samaritan duties to rescue require us to consider the distribution of burdens that duties to rescue generate. This is a feature of other duties as well, including the natural duty of justice.¹⁰¹ For large, ongoing societies, justice requires institutions, and institutions require active support from those within, and non-interference from those without.¹⁰² Furthermore, large-scale institutions usually cannot be established or maintained by a single individual acting alone. Thus, the burdens of establishing and maintaining just institutions must be shared, which entails the further requirement that they be shared fairly. As we have seen, however, this is not enough to support duties that are range-limited or non-discretionary.

If you and I have a natural duty to aid the poor, we each have a duty to do our fair share of the work. If I am unwilling to do my part, you may be required to take up some of the slack I leave. If so, I am blameworthy for imposing unfair burdens on you. However, this does *not* imply that there is some particular way in which I must discharge my duty to aid the poor. The mere fact that the burdens of providing aid must be fairly distributed does not automatically imply anything about how those burdens must be borne. The distribution of burdens raises substantive issues of fairness, but the underlying natural duties remain BAD.

¹⁰¹ The claim that the natural duty of justice is distinct from the natural duty to rescue is controversial. Wellman, as we have seen, argues that political obligation derives from the natural duty of justice, and Wolff (1995) treats the two as interchangeable.

¹⁰² While my focus is on institutional justice, many of the considerations discussed here reproduce themselves at the individual level. If my neighbor and I each have a natural right to hunt in the same woods, we may need to figure out how to distribute access to game. There is a first-order fairness issue here. What is a fair distribution of game or access to game? But maintaining this fair distribution of access might involve bearing some costs. Perhaps we decide to split our catches 50/50 and agree to each show up and do half the work of cleaning the animals and preparing the meat. If I regularly fail to show up for this work but still take half the meat, I am creating a second-order distributive unfairness in the costs of maintaining the fair first-order scheme.

What *can* tie me to particular individuals or institutions through duties that are range-limited and non-discretionary is the receipt of benefits. If just institutions are established in my area, then, to the extent I benefit from them, I have an obligation not to free-ride on them. I owe it to those who produce the benefits to do my part in supporting the institutions that benefit me. Thus, other things equal, I have an obligation to prioritize supporting the local institutions of justice that benefit me over supporting or establishing just institutions elsewhere.

Because of its connection to a distribution of benefits, the natural duty of justice invokes the PoF in ways that the natural duty to rescue does not. Samaritan duties to rescue generate asymmetric relationships. The rescuer bears burdens, while the rescued receives benefits. Thus, a fair relationship among rescuers can only be concerned with the distribution of burdens. Local institutions of justice, however, return benefits to the local actors who support them. Thus, the natural duty of justice can concern itself with the distribution of benefits returned to those who bear the burdens of generating them. From this perspective, the Kant/Waldron duty to create local security is a better candidate than the duty to rescue for grounding an obligation to obey the law since security is a local benefit.

Still, one might wonder whether the beneficiaries of rescues count as free-riders if they fail to subsequently join their rescuer's future rescue efforts (cf. Cullity 2003). Being rescued is obviously a benefit, so, the thought goes, if the mere receipt of benefits is all it takes to trigger fair-play obligations, then they should be triggered by rescues. This would impose an obligation on the rescued to pay back their rescuers. But this misses the point of rescuing. Natural duties to rescue derive from the value of the rescued—she is *entitled* to be rescued. And since she receives only what she is entitled to when she is rescued, she has no special obligation to reciprocate in

kind to those who rescue her. If natural duties do raise considerations of fairness, they should not yield the result that the rescued owe special obligations of reciprocity to their rescuers.¹⁰³

The point I want to emphasize is that duties create roles and sort people into moral relationships. In the relationships created by the duty to rescue and other similar duties to aid, there is a clear assignment of benefits and burdens. Rescuers should bear the burdens of rescues while the rescued should receive the benefits. Questions of distributive fairness can arise within each group, but not between the groups. I explored the fairness obligations that arise among rescuers in chapter four of this dissertation, so here I will only briefly note the similar obligations that can arise among the rescued.

Those who benefit from rescues have an obligation to other rescuees not to take more than their fair share of rescue benefits. If I and two others are drowning, and three life-rings are thrown to us, I have an obligation to take only one (assuming each of us needs only one to survive). Or suppose that I am out of work and in need of career counseling. Assuming there are others in a similar situation, and that the career counselor has limited availability, I have an obligation to other advisees to take no more than my fair share of her time.¹⁰⁴

There is a crucial difference between fairness among rescuers and fairness among the rescued, however. Among rescuers, doing one's part involves contributing enough to some rescue effort or other. One is not bound to any rescue effort in particular. As one of the rescued, however, one's obligation not to free-ride is owed to those who also benefit from the particular rescue effort. When I take up too much of the career counselor's time, I free-ride only on the

¹⁰³ This is not to say that the rescued owe nothing at all. They may owe a debt of gratitude to rescuers, for instance, or might have reason to prioritize rescuing a former rescuer if their positions are later reversed. What is important here, however, is that the rescued have no immediate obligation to reciprocate in kind.

¹⁰⁴ Taking more than my fair share is a kind of active free-riding. By hoarding benefits, I actively free-ride on the restraint shown by others in need who also stand to benefit from career counseling. In public-goods scenarios, by contrast, free-riding is passive; I passively receive a benefit and then refuse to contribute my fair share to the scheme that produces it. I will discuss publicity in section 4.

restraint of those who might schedule a meeting with her. Fairness obligations among beneficiaries are range-limited, but are boundless among rescuers.

This difference is crucial for the natural duty of justice. The groups of those who benefit and who bear burdens can, and often do, overlap. Where they do, fair distributions of benefits are automatically range-limited to those who receive them. Thus, the natural duty of justice creates space for a much fuller application of the PoF than does the natural duty to rescue. As I develop this connection over the next two sections, I argue that, while the PoF comes into play, it, like the natural duty of justice, is altered by their interaction in ways that allow the principles to be mutually reinforcing.

Before proceeding to these arguments, I should address a preliminary worry about range-limitedness. In the case of burden-dumping, I have a duty to all those in a position to perform rescues that I do my part of the total pool of rescues shared globally. If I am not doing my part, those rescuing migrants crossing the Mediterranean have a fairness complaint against me even if, when I do my part, I decide to put out forest fires in the Amazon. We might wonder whether something similar is true for beneficiaries under the natural duty of justice. If I take more than my fair share of the benefits in one scheme, perhaps I can offset this by reducing my consumption of benefits from another scheme, or by working harder to produce more benefits for others elsewhere.

Neither strategy addresses the unfairness of taking more than my fair share of benefits. In the career counselor case, the fairness complaint is not that I take more than my fair share of the total pool of all available benefits, but rather that I take more than my fair share of *this* benefit. I do not make things more fair for others in need of counseling by refusing, say, unemployment benefits, or by working to fund career counseling services on the other side of the world. What is

unfair is that I take more than my fair share of this benefit, and those whom I treat unfairly are those within the same pool of potential beneficiaries. The same reasoning applies to passive free-riding. What is unfair about benefiting from, say, universally available health care without paying my fair share of the health-care tax is not that I receive too many benefits overall. I might receive less than my fair share of benefits overall. If I live in a poor country, rich countries might have cosmopolitan duties to redistribute wealth in ways that would make me much better off. Still, my dodging the health-care tax is unfair to those who are doing their part to support health care in my community.

Paradigm cases of free-riding involve individuals who receive benefits they regard as worth their cost, but avoid contributing for selfish reasons. But not all free-riding is like this. Imagine I have a neighbor who, like me, does not pay the health-care tax, but unlike me, is a saint who makes massive contributions to justice elsewhere. Does she treat anyone unfairly when she refuses to pay the health-care tax? How we answer this question will depend on the role we think efficiency plays in fixing our duties of justice. Here I understand efficiency to be the ratio of justice output to cost input. The larger the ratio, the more efficient the justice-promoting scheme. This notion of efficiency allows us (in principle) to create a rank ordering of justice-promoting schemes in which there may be ties. An intuitive thought is that we are morally required to contribute to the most efficient scheme, in which case the natural duty of justice would be non-discretionary (unless two or more schemes were tied for most efficient). I assume, however, that the bar is not set so high. Above some threshold of efficiency, we have discretion about which scheme or schemes to promote.

As will become clear in the next two sections, the PoF prioritizes contributing to institutions of justice from which one benefits over contributing to those from which one

receives no benefits. But this alone does not tell us much. If the prioritizing is minimal, the duty of justice will almost always outweigh the extra weight given to local institutions by the PoF and require us to make our contributions in places where injustices are most severe and easiest to solve. If the prioritizing is substantial, obligations to local institutions will often outweigh considerations of efficiency.

So far, I have argued that the duties imposed by the natural duty of justice can be modified by the PoF by sorting actions supporting or creating just institutions into low or high priority (subject to efficiency constraints). Actions that provide fair reciprocation for mutual benefits have priority over those that do not. Thus, the combination of the principles solves, or at least mitigates, the particularity problem. However, if the PoF is left unmodified, the worries about its generality that threaten to undermine it as a viable principle of political obligation will also threaten to undermine its viability as a solution to the particularity problem. Thus, we must now see how the natural duty of justice modifies the PoF.

3. The Principle of Fairness and Generality

The PoF, operating in isolation, appeals to the receipt of benefits to justify the imposition of burdens. If the ‘benefits’ one receives from a public-goods scheme are actually detriments, then clearly there is nothing to be repaid. Similarly, if the benefits one receives are small compared to the costs of doing one’s part, they are not worth their cost and no obligations arise. These features of the PoF are the ones that raise worries about generality. If too few people receive benefits that are worth their cost from government actions and programs, the PoF will not generate sufficiently widespread obligations to obey the law to count as a theory of political obligation.

On the view I develop here, however, the PoF does not have to do the work of justifying the imposition of burdens. That work is done by the natural duty of justice. However, the PoF still generates obligations on the basis of benefits received. What is different is that these benefits no longer need to outweigh burdens to generate obligations to reciprocate.¹⁰⁵ This decouples benefits from burdens within the logic of the PoF. This doesn't mean, however, that massive burdens can be justified by tiny benefits. The PoF operates within the scope of the natural duty of justice, and no one supposes that the duty requires massive sacrifices for tiny gains. Rather, it is because of our essential interest in living under conditions of relative justice that we have a natural duty of justice in the first place.

It is, I think, clear how this combination of principles answers the generality worry facing the PoF. Our political obligations are ultimately rooted in the natural duty of justice, which, by hypothesis, applies to everyone. As discussed above, the PoF takes this boundless duty and prioritizes range-limited obligations to aid those who institute and maintain the just institutions from which we all benefit.¹⁰⁶ So long as those under the just institutions benefit at least to some extent, they are, to that extent, obligated to do their part in supporting those institutions.

To illustrate, suppose that contributing to local institutions of justice is above the discretion threshold (discussed in the previous section) in terms of efficiency, but that I benefit very little from those institutions. In this case, I am obligated to bear my full share of the costs, even if they clearly outweigh the benefits I receive. As noted earlier, the PoF does not justify the

¹⁰⁵ It is important to note that (so far) the rest of the PoF is left intact. This means that benefits must be mutually advantageous, public, efficiently produced, and so on. I argue in section 4 that in the present context the publicity requirement should also be dropped.

¹⁰⁶ It is worth noting that the mutual benefit condition of the PoF implies that individuals establishing just institutions in a distant land do not generate fairness obligations in the beneficiaries of those institutions. Of course, if those under the institutions start supporting the institutions themselves they will come to have obligations to each other to maintain them, but not to those who founded them. In this context the duty of justice behaves like the duty to rescue. However, if those in the distant land established just institutions that benefited those who benefited them, then the two sets of institutions together might be considered a single mutually beneficial scheme.

imposition of burdens on this view, but rather justifies prioritizing them over other burdens. Thus, prioritizing the obligations of schemes from which I benefit *to the extent to which I benefit* refers to the moral significance of the prioritizing. If I benefit very little, it is not a big deal if I blow off my obligation and contribute to some other scheme. If I benefit a lot, the priority is quite stringent.¹⁰⁷

The picture as presented so far poses several very difficult measurement problems. For one, we must measure efficiency. This is challenging because justice output to cost input may vary from person to person depending on people's willingness and ability to bear the costs imposed by a scheme. Tradeoffs between the outputs of different kinds of justice-promoting schemes may be difficult or impossible to establish, and the same may be true for different kinds of inputs. Once a rank-ordering is established, principled reasons must be found to justify the placement of a discretion threshold above which one is free to choose which justice-promoting schemes to contribute to. Finally, we must determine how much individuals benefit from local institutions of justice and convert this into a stringency measurement that tells us how important it is for each individual to prioritize contributing to the institutions that benefit her.

While I do not attempt to address these challenges here, I suggest that they can be effectively circumvented by treating the obligation to *prioritize* schemes from which one benefits as a *presumptive* obligation. On the presumptive obligation approach, one is required to do one's part to support justice-promoting schemes from which one benefits unless one can demonstrate that one has discharged one's duty of justice elsewhere. So formulated, the presumption is very weak. If I can demonstrate that I have done enough for justice (above the discretion threshold), I

¹⁰⁷ We might think the cost to others of failing to do my part should play a role as well. If I don't contribute, the local scheme may lose some economy of scale which means burdens for others will increase dramatically. This an important, but it is an efficiency consideration. Thus, it will be factored in earlier in the ranking of schemes by efficiency and the placing of the discretionary threshold.

needn't do anything for justice locally. This presumption is too weak to be theoretically satisfying, but it may be good enough for practical purposes.

In fact, something like the above presumption seems to be present already in the US tax code in the form of income tax exemptions for charitable giving. The list of exemption-granting donations includes more than just contributions to justice-promoting organizations, but the overwhelming focus is on charities that can plausibly be seen as working to redress distributive injustices (IRS 2020). The presumptive obligation framework I have just described provides a justification of these exemptions. Residents of the US have a presumptive obligation to pay their fair share of taxes unless they can show they have done their part (financially) to support justice by giving their money to institutions other than the US government.¹⁰⁸ However, interpreting the PoF as establishing priority is, I think, most satisfying theoretically and it is the interpretation I will assume henceforth. The presumption approach is best seen as a rough-and-ready approximation of the logistically much more daunting priority approach.

The combination of the natural duty of justice with the PoF answers the generality objection to the PoF, but with two caveats. First, those for whom the government provides no gross benefits (as opposed to no net benefit) will have no obligation to support it. This is surely the correct result. That said, it is hard to imagine that genuinely just institutions would be of no benefit at all to those under them. Second, those who benefit only a little from just institutions will only be weakly obligated to support those institutions. One might worry that if most citizens receive only small benefits, the level of enforceable taxation thereby justified will not be enough

¹⁰⁸ The fit here isn't perfect. Charitable contributions in excesses of what one's taxes would have been should cancel out one's taxes entirely. Still, treating charitable giving as tax exempt rather than as counting against one's taxes can be seen as a compromise that attempts to balance the following: the discretionary nature of duties of justice, obligations to contribute to schemes from which one benefits, and the need of governments to secure the revenue they need to function.

to support the full range of government spending. But this is not obviously a drawback. Government spending on programs that are not beneficial would be, quite literally, a waste. What is important to emphasize is that this approach justifies obligations to obey the law more easily than the PoF operating in isolation. Alone, the PoF only justifies obligations to do one's part in schemes that provide net benefits. Here, obligations can arise even when burdens outweigh benefits. So long as gross benefits are considerable, obligations to bear a fair share of the burdens will be stringent independently of costs. The PoF uses benefits to prioritize burdens the imposition of which is justified independently by the natural duty of justice.

This argument leaves room for the idea that supporting local institutions may not fully satisfy our duty to promote justice. Suppose, for instance, that I am obligated to bear burdens of cost X in supporting local just institutions. The natural duty of justice, however, requires me to bear $2X$ costs (efficiently) to support justice overall. Once I do my part locally, I would then be free to contribute to maintaining or establishing justice however I like, wherever I like. Since the PoF no longer applies after the first X costs, the discretion of the natural duty of justice returns to me for the remaining X costs. This, again, strikes me as a strength of the theory. It can support a defeasible obligation to prioritize supporting local institutions without artificially limiting the potential duties of justice we might have elsewhere.

Finally, the present theory can explain obligations to contribute to international efforts to promote justice. If, for instance, it is a matter of justice that we urgently address climate change by reducing emissions, and other countries institute an international scheme to that end, the benefits we receive from these efforts require us to do our part as well. The PoF acting alone might impose similar requirements, but by framing the issue in terms of duties, we can more easily answer objections related to costs. Combating climate change might be costly, especially

in the short term, but we can point out that it is our duty to bear these costs. Then, if asked why we should combat climate change *this way*, we can point out that we have an obligation to reciprocate for the benefits we receive from the mutually beneficial sacrifices of others.

In this and the previous section I have argued for a two-stage MP theory that begins with the natural duty of justice, and then appeals to the PoF. The PoF, working within the scope of the natural duty of justice, answers the specificity problem by arguing that we ought to prioritize supporting institutions from which we receive benefits. Next, when the PoF is operating in conjunction with the natural duty of justice, it no longer needs to do the work of justifying the imposition of costs. This answers the generality worry facing the PoF. Because the theory only needs to consider gross benefits rather than net benefits, one of the main obstacles to the generation of obligations is removed.

4. The Principle of Fairness and Completeness

Modern governments do many things, many of which produce public goods, sometimes just by being available. Support for public health is an example of this type of good. Medical care is a potentially excludable good, but when it is generally available, everyone benefits from its availability. Welfare programs function in a similar way for those who are well-off by providing security against homelessness, joblessness, and so on. Other programs, such as support for the arts, contribute to the public good of a robust and flourishing culture.¹⁰⁹ Some governmental activities and services, however, may not produce public goods, or produce only low-value public goods. Alternatively, some goods provided publicly might be easily convertible into excludable goods (health care, for instance, falls into this category). These services may

¹⁰⁹ Klosko presents welfare and support for the arts as paradigm cases of governmental services that do not produce public goods, though my comments here suggest this is mistaken.

produce substantial goods for all or nearly all citizens, but the goods they produce could be made entirely excludable.

The number and importance of government services that are not public goods, or are not significantly intertwined with public goods, is controversial.¹¹⁰ My aim here is to preempt the worry that many important goods and services provided by government are not public, or not necessarily public. On a standard PoF approach, if a significant number of services fall into this category, the PoF will not meet the comprehensiveness requirement. It will generate obligations to obey a portion of the laws, but the portion will be too small. We might wonder, then, why the PoF is limited to public goods in the first place.

Nozick's critique of the PoF includes the observation that "you may not decide to give me something, for example a book, and then grab money from me to pay for it, even if I have nothing better to spend the money on" (Nozick 1974, 95). For Nozick, this has to do with his notion of private property and the entitlements that come with it. More generally, however, we might think it a virtue of a system that it avoid imposing burdens without first gaining some form of voluntary agreement from those who bear them. Thus, to the extent possible, it is better to distribute excludable and optional¹¹¹ goods according to some principle of voluntary exchange, than to distribute public goods according to a principle like the PoF that (on many interpretations) generates obligations based on the mere receipt of benefits.

In light of Nozick's observation, public goods constitute a special case because they cannot be distributed according to principles that operate on prior agreement. Public goods simply fall on people whether they agree or not. We could respond by denying that the provision

¹¹⁰ Klosko (2019) assumes that many goods and services provided by the government are not public. For compelling argument to the contrary, see Arneson (2013).

¹¹¹ See chapter 1 section 1.1 for an explanation of public goods as excludable and optional.

of public goods produces obligations to reciprocate (except when one explicitly agrees to do so). This, however, creates a perverse incentive against agreement. So long as others support the scheme from which I benefit, I can get the benefit without contributing so long as I don't agree to do my part. If many people behaved this way (and it is sometimes assumed that many people would), public goods could not be reliably provided and essential services such as national defense, provisions for public health, and so on, would collapse. Because the provision of public goods is unstable in this way, and because many people do in fact want them provided, we need a way to determine which public goods schemes we are obligated to support, even if we don't explicitly agree to receive their benefits.

This argument for restricting the PoF to public goods rests on the assumption that, other things equal, it is better to impose burdens only on those who agree to bear them than on those who do not. Without this assumption, there would be no reason why someone should not throw a book at me and then grab payment, so long as the cost to me is fair (according to our background theory of distributive justice), and the book really is a genuine benefit to me despite the cost. Put another way, the PoF, when it is operating alone, must justify the imposition of burdens on those who have not agreed to bear them. If it did not have to bear this justificatory burden, it could apply to the provision of excludable and optional goods just as well as to public ones.

Just as the natural duty of justice relieves the PoF from the need to weigh burdens against benefits, so it relieves the PoF from justifying the imposition of burdens no one agreed to bear. We have a duty to promote justice whether it is provided through public goods, or through private goods available to all on fair terms.

5. Reciprocity

Reciprocity has to do with payback, and the question for the PoF is, what counts? The most common answer is that appropriate payback comes in the form of doing one's part within the scheme from which one benefits. Part of the scheme is its division of benefits and burdens, so shirking one's burdens, as assigned by the scheme, while still receiving its benefits, would be a failure to reciprocate. However, intuitions can go different directions on this point. One might think that reciprocity can take any form so long as it offsets others' burdens at least as much as doing one's part as assigned by the scheme. Rather than participate in nighttime patrols with my neighborhood watch, I can bake cakes for those on duty.

The idea that reciprocity in the PoF can take any form might seem odd for a couple of reasons. For one, to take the previous example, cakes do not directly contribute to the neighborhood watch scheme. If enough people decided to bake cakes instead of participating in patrols, the scheme would collapse. When Rawls says that those who have submitted to constraints on their liberty have a right to similar submission from those who benefit (Rawls 1999, 96), he presumably means that those who have submitted to the particular burdens imposed by the scheme have a right against others that they submit to those same burdens. It would not be good enough for others to submit to an equivalent quantity of burdens that produce offsetting benefits for those already doing their part.

But the stability consideration appears merely practical. If enough others are doing their part, are likely to continue doing their part, and I know this, why must I submit to the particular burdens assigned to me by the scheme? This is a version of the particularity objection discussed above. Even if I am tied to this scheme, why am I tied to its particular burdens?

As noted, intuitions go different ways on this question, and most lean toward the conclusion that, under the PoF, reciprocity is owed as bearing the burdens assigned by the scheme. While my intuitions also lean in this direction, I do not have the space to provide a detailed analysis here. However, since what is at stake is a version of the particularity problem, some comments are in order.

I have defended a non-voluntarist version of the PoF according to which mere receipt of benefits generates obligations to do one's part. On this interpretation, one individual or group can unilaterally generate obligations for others by creating the right sort of public-goods scheme. If so, why not think that those who benefit have similar unilateral discretion to decide how they will reciprocate? This way, a kind of symmetry is maintained between the groups—both have an opportunity to choose how they benefit others. Voluntarist interpretations of the PoF that require benefits to be voluntarily accepted before obligations arise can say that acceptance includes accepting the terms of reciprocity. It is not obvious that something similar can be said for non-voluntarist versions.

A Wellman-style intuition about the value of discretion might be surfacing here. Those who establish a scheme have a kind of discretion that other beneficiaries do not. They are morally free to abandon the scheme before it stabilizes, or to decide against establishing any scheme at all. Even in the context of this chapter, establishing a mutually beneficial scheme that promotes justice limits others' freedom in how they promote justice. They face a more strictly prioritized set of options than those who establish the scheme.

It is worth emphasizing again that we do not have a general interest in discretion. Martha Nussbaum and Will Kymlicka have both forcefully argued that we do not have a general interest in freedom or liberty as such. Rather, we have interests in *important* freedoms and liberties

(Nussbaum 2003; Kymlicka 2002). Kymlicka gives (a version of) the following example. Consider traffic laws and freedom of speech, and suppose that I drive almost every day, but almost never say things that are controversial. Under these circumstances, it would be a liberty windfall for me if traffic laws were repealed while freedom of speech was restricted. I would still be free to say everything I want to, and I would additionally be free to park in disabled spots, drive on the left side of the road, turn left on red, misuse my turn signals, drive as fast or as slow as I like, and so on. But obviously all these liberties are unimportant. What's more, they undermine an important interest I have in access to safe and reliable transportation (Kymlicka 2002, 143).¹¹² These arguments apply equally to discretion, so the question is whether we have an important interest in deciding for ourselves how we want to reciprocate for the benefits we receive from public-goods schemes that promote justice.¹¹³

We might think that our interest in contributing to justice as we see fit is quite strong. A core claim of the liberal tradition is that we have a deep interest in forming and pursuing a conception of the good, and a conception of justice is surely a crucial part of it. Limiting a person's ability to act on her own conception of justice might therefore be seen as a serious limitation of her autonomy. At the same time, of course, well-designed institutions can be enabling rather than limiting. Contributing to institutions that provide, at a reasonable cost, something like Rawls's primary goods might be like obeying traffic laws. Provisionally, I suggest that this is so, and that in these cases we do not have a strong intrinsic interest in having complete discretion to pursue justice as we see fit. I say 'complete discretion' because the extent

¹¹² Also, some liberties are important even if I don't exercise them, and freedom of speech falls into this category.

¹¹³ The broader question is about our interests in reciprocating to public-goods schemes generally. Given the range of possible schemes, the interests involved will vary. Additionally, some people might be unable to bear the burdens that directly support a scheme. If I am sickly and bedridden, I won't be able to go on nighttime patrols with my neighborhood watch. But schemes and their terms are not set in stone. They may be bundled together with other schemes, and a variety of contributions can be recognized as forms of reciprocity.

to which political obligations limit our discretion is unclear. If the natural duty of justice demands significantly more of us than we pay in taxes, for instance, the obligation to pay taxes might be a relatively minor limitation on our discretion. Similarly, if a presumption approach is taken at the level of policy design, those who want to exercise discretion may.

The objection I am considering is that establishing a scheme objectionably limits the discretion (or freedom or liberty) of beneficiaries. This objection can be taken in several directions, one of which I have just considered. The argument was a distributive one about the intrinsic value of discretion. There are two more I will consider, one distributive and one relational. The distributive concern has to do with equality. According to this objection, the asymmetry of discretion between founders and beneficiaries is unfair because those who establish a scheme have the option to establish it or not while those who benefit are constrained to do their part (unless some excusing condition applies). Equality is intrinsically valuable, but, the objection says, the PoF endorses inequality. Relationally, we might think of asymmetry as a kind of exploitation or subjugation. Those who establish a scheme gain power over beneficiaries in a way that objectionably subjects them to their will. I will consider each objection in turn.

One way to reject the inequality objection is to deny the intrinsic value of equality. There are strong reasons to think that equality is at best instrumentally valuable (cf. Arneson 2015), and it is difficult to see what its value might be in this case. Income inequality can be objectionable if it leads to social instability; unequal access to positions of advantage (whether formal or substantive) can be objectionable if it labels some as second-class citizens. Inequalities in opportunities to establish just institutions do not lead to these particular negative outcomes. Properly functioning institutions of justice produce stability and security no matter who establishes them. Similarly, even if some (constitutional framers, for instance) have an

opportunity to shape social institutions that others lack, no one in a well-functioning democratic society is relegated to second-class status. Instead, everyone has comparable opportunities to participate in and influence the existing institutional structure. Real democracies do not live up to this standard, but the problem is not that the PoF unequally distributes opportunities to establish institutions of justice. The problem is that the current systems lead to pernicious outcomes. Distributing some goods or opportunities more equally might be a way to avoid these outcomes, but if not, the mere fact that they are more equal is not a reason to favor them.

Even if we grant the intrinsic importance of equality, it is not obvious that the PoF leads to objectionable inequalities of opportunity. Before anyone establishes a scheme, everyone is equally free to do so, at least in a negative sense. In this way, there is no background unfairness to one person or group taking an opportunity and thereby preempting others. Of course, unfairness might infect the scheme if background conditions are unfair. Starting a particular scheme might take a certain amount of leisure time or wealth, and if these are unfairly distributed to begin with, then one person's initiating a scheme could represent an unfair advantage on her part. In this case, though, the unfairness does not derive from the fact that one person or group initiated a scheme while others did not, but from the unfair initial distribution of resources or opportunities.

Additionally, as noted above, one person's or group's initiation of a scheme might, all things considered, expand others' opportunities to initiate or contribute to schemes that interest them. In this case, those bound to others' schemes will not face exactly the same options as those faced by the people who started the schemes, but they will face options that are at least as good. If we think some options are more valuable than others, we need to make sure that each person's bundle of options contains good options by quality, not just many options by quantity.

Regardless, if our concern is equality, it is enough that everyone's opportunities are equally good, even if their option sets are different.

Turning to the relational objection, worries about exploitation and subjugation might stem from either the hierarchical structure of a public-goods scheme (if it has one), or from disproportionate benefits received by those who institute the scheme. In both cases, the worries are substantially addressed by the requirement that the scheme be fair according to whatever theory of distributive justice we accept. A scheme might be hierarchical, but it will not allow those in charge to take advantage of those under them. To the extent that a scheme is unjust, obligations to obey it do not arise, or are overridden by considerations of justice.¹¹⁴

Finally, we might simply say that it would be nice to be free to choose how one supports justice. This may be true, but it is not enough to show that we have a morally significant interest in this kind of discretion. What's more, for many living in wealthy countries, doing their part to support local institutions of justice may not exhaust their natural duty to promote justice, and certainly will not exhaust their ability to promote it. People in these categories have full discretion to (efficiently) promote whatever causes they find most compelling or urgent.

Chapter 5, in full or in part, is currently being prepared for publication of the material. Finley, Aaron. The dissertation author was the primary author of this material.

¹¹⁴ In his discussion of the PoF, Rawls refers to institutions that are "reasonably just in view of the circumstances" (Rawls 1999, 96). Simmons argues in response that the justness condition can be dropped altogether (Simmons 1979, 109-114). I suggest that, so long as a scheme is producing benefits that are worth their cost, costs imposed by the scheme should be discounted in proportion to the degree of injustice suffered by individual beneficiaries, which includes efforts to obstruct improving or replacing unjust schemes with better ones. Thus, while the PoF does allow some degree of exploitation and subjugation, this is true of every theory that says some injustice must be tolerated given the world as it is.

References

- Arneson, Richard. 1989. "Equality and Equal Opportunity for Welfare." *Philosophical Studies*, 56 (1): 77–93.
- Arneson, Richard. 2013. "Paternalism and the Principle of Fairness." In C. Coons and M. Weber (Eds.), *Paternalism: Theory and Practice*, 134–156. Cambridge University Press.
- Arneson, Richard. 2015. "Why Not Capitalism?" In A. Kaufman (Ed.), *Distributive Justice and Access to Advantage: G.A. Cohen's Egalitarianism*, 207-234. Cambridge: Cambridge University Press.
- Arneson, Richard. 2016. "Does Fairness Require a Multidimensional Approach?" In M.D. Adler and M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*, 588–614. New York: Oxford University Press.
- Bykvist, Krister. 2016. "Preference-Based Views of Well-Being." In M.D. Adler and M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*, 321–346. New York: Oxford University Press.
- Cullity, Garrett. 2003. "Beneficence, Rights and Citizenship." *Australian Journal of Human Rights*, 9 (2): 85–105.
- Huemer, Michael. 2013. *The Problem of Political Authority*. Palgrave MacMillian.
- IRS. 2020. "Charitable contribution deductions." Retrieved 5/3/2021, from <https://www.irs.gov/charities-non-profits/charitable-organizations/charitable-contribution-deductions>
- Klosko, George. 2004. *The Principle of Fairness and Political Obligation*. Rowman & Little-Field.
- Klosko, George. 2019. *Why Should We Obey the Law?*. Medford: Polity Press.
- Kymlicka, William. 2002. *Contemporary Political Philosophy: An Introduction*. Oxford: Oxford University Press.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Cambridge: Blackwell.
- Nussbaum, Martha. 2003. "Capabilities as Fundamental Entitlements: Sen and Social Justice." *Feminist Economics*, 9 (2–3): 33–59.
- Rawls, John. 1999. *A Theory of Justice, Revised Edition*. Cambridge: Harvard University Press.
- Simmons, A. John. 1979. *Moral Principles and Political Obligations*. Princeton University Press.

- Simmons, A. John. 2001. "Fair Play and Political Obligation: Twenty Years Later." In *Justification and Legitimacy: Essays on Rights and Obligations*, A.J. Simmons (Auth.), 27–42. Cambridge University Press.
- Waldron, Jeremy. 1993. "Special Ties and Natural Duties." *Philosophy & Public Affairs*, 22 (1): 3–30.
- Wellman, Christopher, and A. John Simmons. 2005. *Is there a duty to obey the law?*. Cambridge: Cambridge University Press.
- Wellman, Christopher. 2001. "Toward a Liberal Theory of Political Obligation." *Ethics*, 111 (4): 735–759.
- Wolff, Jonathan. 1995. "Pluralistic Models of Political Obligation." *Philosophica*, 56 (2): 7–27.