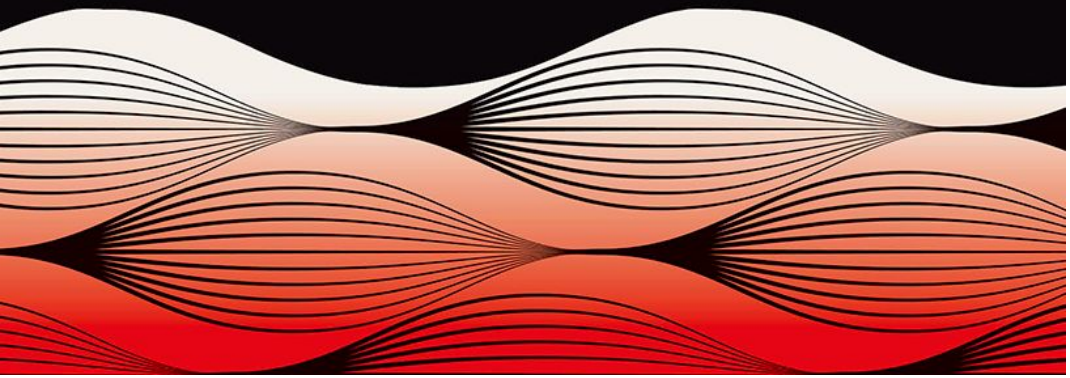


OXFORD

oxford studies in
POLITICAL PHILOSOPHY

Volume 7



Oxford Studies in Political Philosophy

Oxford Studies in Political Philosophy

Volume 7

Edited by

DAVID SOBEL, PETER VALLENTYNE,
AND STEVEN WALL

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© the several contributors 2021

The moral rights of the authors have been asserted

First Edition published in 2021

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019947968

ISBN 978-0-19-289748-0

DOI: 10.1093/oso/9780192897480.001.0001

Printed and bound in the UK by
TJ Books Limited

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Acknowledgments

Almost all of the papers in this volume were presented at the Workshop for Oxford Studies in Political Philosophy, which took place at Syracuse University in 2019. We are grateful to the authors, and to the commentators at that event, for making it a stimulating and productive exchange of ideas. We are also very grateful, as always, to the members of our Editorial Board who provided the helpful refereeing of all the papers. In addition, we would like to thank all those who helped put on the event at Syracuse, including Roberta Hennigan and Lisa Farnsworth. Also thanks for help with the conference to Teresa Bruno and, again, for her fantastic conference artwork, Nikki Fortier.

Special congratulations to Helen Frowe. Her paper in this volume, “The Limited Use View of the Duty to Save,” was awarded the Sanders Prize in Political Philosophy in 2019.

Thanks to Syracuse’s Dept of Philosophy and the Center for the Philosophy of Freedom at the University of Arizona for providing funding for the workshop.

Finally, we want to take a moment to note that the above-mentioned conference and this volume mark the end of Peter Vallentyne’s years as co-editor of *Oxford Studies in Political Philosophy*. The three of us, Peter, Steve, and I, started up OSPP together and Peter’s experience, knowledge, energy, and organization have been crucial in making OSPP better in every way. We will miss him playing this role with us.

Contents

<i>List of Contributors</i>	ix
1. The Philosophy of Work <i>Kwame Anthony Appiah</i>	1
2. Political Epistemology and Social Critique <i>Sally Haslanger</i>	23
3. The Limited Use View of the Duty to Save <i>Helen Frowe</i>	66
4. Politics, Truth, and Respect <i>Jonathan Turner</i>	100
5. Perfectionist Duties <i>Collis Tahzib</i>	124
6. When Public Reason Falls Silent: Liberal Democratic Justification versus the Administrative State <i>Brian Kogelmann and Stephen G. W. Stich</i>	161
7. An Autonomy-Based Argument for Democracy <i>James Lindley Wilson</i>	194
8. Why Strict Compliance? <i>Simon Căbulea May</i>	227
<i>Index</i>	265

List of Contributors

Kwame Anthony Appiah, New York University

Helen Frowe, University of Stockholm

Sally Haslanger, Massachusetts Institute of Technology

Brian Kogelmann, University of Maryland

Simon Căbulea May, Florida State University

Stephen G. W. Stich, Yale Law School

Collis Tahzib, University of Oxford

Jonathan Turner, University of Oxford

James Lindley Wilson, University of Chicago

1

The Philosophy of Work

Kwame Anthony Appiah

My aims in this chapter are modest: I want to suggest that there is a field that has not received sufficient attention in recent philosophical writing. I will call that field “the philosophy of work.” And there are, I think, at least three main philosophical clusters of issues that deserve more study. The first cluster is a set of questions in social ontology. What is work? A job? A profession? A vocation? I think the answer to these questions requires the sort of genealogical investigation that is often required in social ontology. How, for example, did the social world and our concepts develop in a dialectical relation with one another to produce the contemporary idea of the job?

A second cluster of problems is ethical. How does work or a job fit into the good life? At the heart of the issues raised by this question, I am going to argue, is a hard problem. This problem is, in essence, that work has come to matter in a series of interdependent dimensions of social and individual life, and that it is not evident how, as technology and society develop moving forward, we can easily construct new forms of social life that will satisfy human beings in all those dimensions in the way that the best jobs did and do. The rise of the robot and of AI both eliminate and reshape jobs, as we know, in ways that bring costs as well as benefits for human flourishing. So does the globalization of production and distribution of goods and services. These are issues on which many popular writers on the “Fourth Industrial Revolution” and on globalization have, of course, written for some time.¹ But what’s lacking, I claim, is serious organized reflection on the normative issues raised by these challenges.

¹ A typical survey can be found in (Schwab, 2016), where the Executive Director of the World Economic Forum draws the problems to the attention of global business and political leaders.

We need solutions; but we also need to decide whose responsibility it is to find and to shape them.

And that leads to the third cluster of concerns, which is of especial importance for the readers of this book: How should law and other sources of normative authority be configured to allow work to contribute to the flourishing of workers, and how should the opportunities and rewards of work be shared?²

1.

Most healthy adults today spend five days of the week for much of their lives doing what we call “a job.” They may be in factories or offices, or in cars or buses, trains or planes; working in or from their own home or in someone else’s, or in a hotel, restaurant, bar, factory, warehouse, hospital, school, college, or military base. The U.S. Bureau of Labor Statistics declares: “All workers are classified into one of 867 detailed occupations according to their occupational definition.” I find myself, like some of you, at “25-1126 Philosophy and Religion Teachers, Postsecondary.”

Because people typically spend eight hours or more of each weekday at work and another eight or so hours sleeping, it is where they spend about a third of their waking hours. Work is also, as a result, the site of a great number of our relationships. And if we’re lucky, our job is not just a source of a decent income; we will think what we do worth doing. We can also hope to be esteemed for achievements at work. So, we can take pride in what we do, and it may be a source, therefore, of self-esteem. These are possibilities for work of every kind: for the supermarket bagger with Down’s Syndrome, whose cheerfulness is appreciated—and which she knows is appreciated—by many of the shoppers she sees regularly as well as for the litigator whose skills are recognized in awards from her peers; for the elementary-school teacher who watches her protégés grow into successful young men and women as well as for the Nobel laureate in economics or medicine.

² I’m grateful to a reader for this book whose comments suggested to me I should lay out a map of the territory here at the start.

Work can also be a source of identity. It may give you a profession, conferring on you the status of an auto mechanic, a beautician, a journalist, a nurse, a lawyer, or a teacher; often, along with wealth, education, and connections, your job helps fix your social class.

Having started working in our late teens or early twenties, we're likely to continue working for another half-century or more. And, in retirement, when we're no longer doing paid work—or, at any rate, doing much less of it—we may suffer the loss of the sense of purpose that an occupation once gave us and feel nostalgia for the daily rhythms of our job.

Given this centrality of work in our lives, I think it odd how little space it takes up in contemporary ethics and liberal political philosophy. This silence about work echoes our long silence about the family, which feminist philosophy has remedied. It will be important to keep track of gender and family in thinking about work as well.

We do, of course, regularly discuss some of the *proceeds* of work: income and wealth and, more recently, esteem or respect. We think about the allocation of these in political philosophy when we discuss distributive justice and equality. We recognize in ethics and moral philosophy generally that the character of our relationships matters. And this thought has come to be central to more recent thinking about political equality, too. But the focus of our interest is more likely to be on relations with our fellow citizens generally—on what it means to treat each other as equals, for example—than on how we interact with others in the workplace in particular.

Indeed, at least in the United States, we take it for granted that at work we are *not* equals. Most work is organized hierarchically: there are managers, bosses, deans, and CEOs. We have spent a great deal of time in political philosophy reflecting on how the state and its agents can derive the authority to command the citizen. Only recently, I think it is fair to say, have philosophers begun to take with full seriousness questions about democracy in the workplace.

There are exceptions to the relative silence about work as a philosophical problem. Elizabeth Anderson's *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)* (Anderson, 2017) is an eminent, exemplary, and excellent recent example here. Axel Honneth

(1996) has connected work with important issues of recognition in *The Struggle for Recognition: The Moral Grammar of Social Conflicts* and some of his writings since. He is following in the footsteps of Marx and the Marxists, who thought a great deal about work, in ways shaped by their debt to Hegel.

In the 1950s, Hannah Arendt, writing in *The Human Condition*, distinguished *labor*—making or acquiring food and shelter, and doing all the other natural things necessary to sustain our biological life—from *work*, which is the artificial shaping of the world to make products. (She had a final category, *action*, which involved what we do with one another.) And her book, which first appeared in 1958, already addresses the significance of automation. We are, she wrote, “a society of laborers which is about to be liberated from the fetters of labor, and this society does no longer know of those other higher and more meaningful activities for the sake of which this freedom would deserve to be won” (Arendt, 1998, p. 4).

The general consensus since then, however, has essentially been that Arendt need not have worried. Our conception of what is valuable in human life has been so profoundly formed by the place of work for people today that many just assume we will find new ways of making work, even if the biological needs that Arendt’s *labor* was meant to meet can all be met by intelligent machines.

2.

The jobs created by the Industrial Revolution did at least four important things. First, of course, they produced goods in larger quantities with increasing efficiency. A second thing they did was to provide employees and shareholders with income. They built on the genius of capitalism for taking one person’s savings and combining them with the industry and ideas of others to produce an income for them all.

A third important consequence was the creation of new forms of community. Trade unions come with union picnics, and factories may have sports teams and Christmas parties (McIntosh, 2011). At work itself, too, at least in the best of jobs, one’s product is the result of

rewarding social processes, the combined effect of the coordinated interactions of human beings collaborating, working together.

The final, fourth, contribution was that, if you were lucky, your work was a source of significance. The Working Men's Associations of nineteenth-century Britain were reflections of a growing pride in manual labor. People came to appreciate that the goods they helped make were important to their country and its people and were often valued by others at home and around the world. Writing about nineteenth-century trade unions, E. P. Thompson says (in his classic book on *The Making of the English Working Class*), "Social and moral criteria—subsistence, self-respect, pride in certain standards of workmanship, customary rewards for different grades of skill—... are as prominent in early trade union disputes as strictly 'economic' arguments" (Thompson, 1966, p. 236).

William Blake may have seen in the factories of the Industrial Revolution only "dark, Satanic mills": their denizens increasingly saw the work they did as a source of pride, identity, and meaning. And the associations, unions, and clubs they formed—in which people who worked together, played together—came to perform a role not just in their political lives but in their social lives as well.

By the time the Universal Declaration of Human Rights was formulated, just after the end of the Second World War, Article 23 guaranteed "the right to work, to free employment, to just and favorable conditions of work and to protection against unemployment." But Article 24 immediately added that, "Everyone has the right to rest and leisure, including reasonable limitation of working hours and periodic holidays with pay."

The working class was following here a path set earlier by the middle classes. Through the course of the nineteenth century romanticism encouraged an ideal of self-development, which we see in Matthew Arnold's condemnation of Philistinism in *Culture and Anarchy* (1869), and in John Stuart Mill's celebration, in chapter 3 of *On Liberty* (1859), of "individuality as one of the elements of well-being." Playing and listening to music, reading literature, writing and reciting poetry, painting, sculpting, visiting art museums, learning history and social science, even following the sciences of their day: all came to be part of what was expected of an educated middle-class man or woman. This is what Mill meant when he talked about individual development.

The Germans called this form of cultivation “Bildung,” and European societies had a growing class they later called the Bildungsbürgertum, the educated bourgeoisie.³ One of the central questions that received the attention of German philosophers, beginning in the early nineteenth century, was what they called die Soziale Frage, *the* social question: at its heart was the welfare of the new working class, created by industrialization, urbanization, and an expanding population. Eventually Bildung came to be part of the story; for Bildung was now a component of a normal human life. It was not just for a leisured aristocracy or for the free time of a middle class.

And so, starting at the end of the nineteenth century, in many places in Europe and North America, new institutions were created to extend the benefits of Bildung to workers. Beginning in the 1880s, the settlement house movement in Britain and the United States moved middle-class “settlers” in alongside working-class families, in part so that the former could share their “culture” with the latter. In 1899, Ruskin College was founded in Oxford to offer a tertiary education to workingmen who did not have access to Oxford University. This democratization of learning became one of the founding aims of the British Broadcasting Corporation in 1922, under the leadership of John Reith. Less than half a century later, Lord Reith’s ideals were reflected in the founding of the American Corporation for Public Broadcasting, whose stated purpose “is to provide programs and services that inform, educate, enlighten, and enrich the public and help inform civil discourse essential to American society.”⁴

A modern democracy gave the responsibility of choosing who would govern to the people, but the people needed an education if they were to perform this great responsibility well. The vast expansion of higher education after the Second World War, in the United States (and elsewhere), accelerated here by the G. I. Bill, was guided, in part, by the same thought. It was also true, however, that the more lucrative opportunities in the modern workplace were going to those with a college education.

³ The verb *bilden* can mean simply shaping or making; like the French word *formation*, Bildung connects the idea of shaping a person with the idea of a preparation for her life.

⁴ <https://www.cpb.org/aboutcpb/goals/goalsandobjectives>.

What the Bill opened up was the possibility of a huge increase in the numbers of people receiving a liberal education: an education fit for free men and, coming close behind them, women. Other societies have followed. And there were three different purposes woven together in the idea of modern education: as a source of enrichment for people in their private lives; as a preparation for civic responsibility; and, last, though, no doubt, not least, as a preparation for the world of work.

3.

One of the basic social and economic challenges of our time, then, is to find ways of involving people in meaningful activity while, at the same time, distributing the social product fairly, giving everyone a satisfactory income, and producing the goods and services we need. With or without work, we need new ways of providing the four important things that I said good jobs did in industrial society. This is an intellectual and imaginative challenge as much as an institutional one. I am going to call it (with apologies to my friends in the philosophy of mind) the “hard problem.” That we can only solve it properly today if we do so in ways that are ecologically sustainable only adds to its difficulty.

This hard problem, then, is to find ways to produce the goods and services we need, while providing people with income, sociability, and significance. And a major issue is whether we do this by changing the ways we construct and provide jobs—the route of reimagining work—or by meeting these needs for many or most people without their having jobs, as a self-styled “post-work” movement is suggesting. And the thought there is not just that work might disappear for many but that the concept of work is an obstacle to progress.

Technological change means that fewer and fewer people are needed to produce the same quantity of goods and services. The result is that there are many people whose only income comes from the state and private philanthropy or from jobs that lack the satisfactions—in income, meaning, and sociability—that once secured the status of the industrial working class. If they have left the labor pool altogether and are no longer seeking employment, it is not just because there are literally no jobs:

rather, they have given up on finding a job for which they are qualified and that can be a source of self-respect, or, if not, of an income large enough to make up for the fact that the work itself is not a source of self-respect.

Leaving the labor pool means, of course, that they no longer participate in the community of the workplace. But many of the new jobs don't have much scope for sociability either. With people now able to telecommute, even those who *do* have employment may not gain the experience of community from their work. If you work as an Uber or Lyft driver, or in many other occupations in the "gig economy," your assignments are organized without ever bringing you together with others who are doing the same job for the same company.

Worse, many modern people are doing what the anthropologist David Graeber has dubbed a "bullshit job," defined as "one so completely pointless, unnecessary, or pernicious that even the employee cannot justify its existence" (Graeber, 2018a, p. 3). He cites the result of a British YouGov poll that asked, "Does your job 'make a meaningful contribution to the world?'"

Astonishingly, more than a third—37 percent—said they believed that it did not (whereas 50 percent said it did, and 13 percent were uncertain). (Graeber, 2018a, p. xxiv)⁵

Graeber claims that people often conflate "bullshit jobs" with something else... what he calls "shit jobs."

...but [he says] they're not the same thing. Bad jobs are bad because they're hard or they have terrible conditions or the pay sucks, but often these jobs are very useful. (Graeber, 2018b)

⁵ Here are the five major categories of such occupations: flunkies, whose job it is to make other people look good; goons, like telemarketers, whose job is only necessary because there are other people like them; duct tapers, who clean up the results of flaws of institutional design, exemplified, as Graeber told the *Daily Beast*, by the "poor guy at my university whose entire job seemed to be apologizing for why the carpenter could not come and fix the bookshelves in my office"; box-tickers, who do what you'd expect; and, last but by no means least, taskmasters, who "typically provide unnecessary supervision" (Graeber, 2018c).

Both kinds of jobs raise ethical problems, of course. Even if you recognize your job is useful, there's no guarantee that it will contribute to your satisfaction if the pay or working conditions are awful. Increasingly, then, one source of meaning in human lives—the job, the career, and its sociability and its achievements—is going away. And, though this problem has developed first in the industrialized democracies, it will surely eventually spread everywhere.

It is certainly good that machines can be turned to doing and making useful things that it is no fun for people to make or to do. Where possible, what Graeber calls “shit jobs” need to be eliminated (perhaps by improving pay and conditions or by making them easier with robots or AI). It may also be good when the efficiency of production grows, in the sense that it takes fewer and fewer people to make things; though we should give a moment's thought to the possibility that there are automatable tasks that human beings might enjoy doing and receiving the results of.

But our automated economy still makes the things that were at the core of production in the old economy; indeed, we are making more and better things. That, though, can leave income, sociability, and significance unattended to. You could solve the problem of the disappearance of the wage by establishing a basic income guaranteed to all citizens. But that, too, wouldn't help you with the loss of community and the loss of meaning. Conversely, in a society like ours where no one can provide for their basic needs without money or expensive land, a life of sociability and meaning without access to an income is no longer possible. That is why the problem is hard.

4.

There are three recognizably philosophical tasks here, as I said at the start. The first is to explore the concepts of work and job. That project, though in a sense a matter of conceptual analysis, is not, of course, a priori. For work develops along with technologies and institutions, so the inquiry is in part historical because work and the concept of work develop together. Ian Hacking remarks, in the second chapter of *Historical Ontology*, that “Foucault's books are mostly about practices

and how they affect and are affected by the talk in which we embed them. The upshot is less a fascination with words than with people and institutions, with what we do for people and to people” (Hacking, 2002, p. 47). I remember, as an undergraduate, hearing Hacking introduce Foucault’s methods at the Moral Science’s Club at Cambridge. It took me a while to know what to do with those ideas. But I find this basic thought is now an essential philosophical tool.

Once we have understood what work is, that question in social ontology, there is, next, that ethical inquiry I sketched in my introduction: how does work fit into making a good life, advance eudaimonia, help humans flourish? Here, too, the inquiry strikes me as necessarily historical; but it also requires us to draw on our own social experience and on reports—in history, sociology, anthropology, and imaginative literature—of the experiences of others. I don’t mean that the conceptual inquiry is sharply bounded from the ethical inquiry: our understanding of the meaning and the value of work develops historically with the economy, with the institutions and technologies we engage in our work. And the conceptual inquiry is already a normative inquiry, as we have seen, because you cannot understand what a job is without understanding the idea of a good job.

Which brings us to that third set of questions in political and social philosophy: How should work be constrained or constructed by law and other social norms, and how should opportunities and rewards for work be distributed? We have models for thinking about these issues, of course. One is the Rawlsian program, in which we ask how we would respond to these questions if we didn’t know what opportunities we ourselves would have. But the question only arises for Rawls because society and the state are immense and valuable cooperative enterprises whose benefits and burdens must be fairly shared. The basic structure of society—the family, the law, the economy, “the main political and social institutions and the way they fit together as one scheme of cooperation,” as he put it once—must be (as he also put it) “a fair system of social cooperation over time from one generation to the next” (Rawls, 2001, pp. 4, 5). Now Rawls’s program is offered as a contribution to ideal theory. It is worked out for an “ordered society,” whose members and whose institutions are known by all to meet two conditions: they have a

shared commitment to an ideal of justice and their institutions more or less realize it.

Much can be learned by asking the questions Rawls's way.⁶ But there is also a great deal to be learned from an approach I associate with that most philosophical of economists, Amartya Sen. We don't begin with a picture of a just society, not because that picture idealizes too much, but because it misunderstands the epistemology of our moral knowledge about politics. The general point, which Sen has rightly made central to his thinking, is that you can judge social option A better than social option B without starting with a view of the best society and asking whether having A or having B brings you closer to it... just as you can tell that a Rembrandt is better than a Ruysdael without any idea of what the best painting would look like. I think this point, though simple, is a deep and important one. You don't need to know what the heavens are like to know which way is up.

This insight fits with another. Our collective moral learning doesn't require the development of a picture of an ideal society. It starts most often with the rejection of some current actual practice or structure, which we come to see as wrong. You learn to be in favor of equality by noticing what is wrong with the unequal treatment of blacks, or women, or working-class or lower-caste or LGBTQ people. You learn to be in favor of freedom by seeing what is wrong in the life of serfs or the enslaved or of women in purdah (Patterson, 1992).

So, rather than invoking ideal societies, I'd like to ask whether we can move our actual norms, our laws and other institutions, toward the provision for everyone of the resources for a more dignified human life. This is a question that arises from within a society that is up and running. And the critique of current institutions and practices develops because we discover through what Mill called "experiments of living" that features of our current life damage or enhance the possibilities for human flourishing. We start, for example, with gender norms as they are and discover that they are disabling for trans people, and so need revision. No need to think abstractly about the biological significance

⁶ There is a longer discussion of Rawlsian ideal theory in chapter 3 of (Appiah, 2017).

of sexual difference and imagine without presuppositions—painting, as it were, on an empty canvas—what would be the best way of developing a set of ideas and practices around gender.

Because what makes a life of dignity can depend on local cultural understandings, the idea of a dignified life is not external to social arrangements. It is not, that is, something we bring to our question—how to remake the world to enable a dignified life for everyone—from the outside. We can only ask these questions about equality and dignity from within a society and its social understandings. Even when we ask them about another society—as we may—we do so by bringing our understandings of equality and dignity and seeing whether and, if so, how they are expressed elsewhere. And both what is dignified and how to relate as equals are matters of ongoing ethical evolution. In that historical development, there is a kind of dialectical relationship between institutional and technological change and normative understandings, of the sort that is evident in the changing conceptions of what it is for work to be rewarding and how it fits into the project of making a life.

5.

The Luddites were convinced that the mechanization of existing forms of labor would destroy jobs. And so it did, of course. But it also created them. The economist Riccardo Zago has taught me about the economic mechanism here in a variety of cases. Take, first, the mechanization of agriculture, which has certainly reduced the number of agricultural workers. It also lowered food prices and so increased demand, however, creating new jobs in the transportation, distribution, and preparation of foods. As Zago points out, “it is not a coincidence that the Meatpacking districts of New York” and other cities “hugely developed in the same periods in which mechanization occurred” (Zago, 2019). Similarly, the more recent spread of ATMs displaced bank tellers, but increased the profits of banks, allowing them to open more branches, where, as you may have noticed, they hired more workers to do managerial tasks and customer care, jobs where people still have a comparative advantage over machines (Zago, 2019, citing Bessen, 2016).

One current problem in the United States, then, is not so much a consequence of a net loss of jobs as of the human costs associated with the transition from one regime of jobs to another. The sorts of adjustments that occurred in the mechanization of agriculture and the spread of the ATM take time. The literature in labor economics suggests the recent round of job displacements—as a result of automation and of the transfer of jobs to lower-cost labor markets elsewhere—has been accompanied by very slow improvements in employment (Zago, 2019, citing Autor et al., 2013 and Acemoglu & Restrepo, 2017). And, by an improvement, I mean the replacement of lost jobs with ones that are better.

New jobs in an automating economy are usually going to require new skills. Finding or training workers with these skills can take time. A displaced worker's initial value in the new labor market may be lower than her value in the old one; it may take her time or money to acquire the necessary human capital. One source of the growth of high-school education in the early twentieth century in the United States was the government's recognition that the children of the workers displaced from pre-mechanical agriculture needed preparation for new forms of employment (Zago, 2019, citing Goldin & Katz, 2018). And, finally, the new jobs may be in new places, and someone has to bear these costs of internal migration.

These traditional difficulties seem to be accentuated in the current economy by four things. First, progress in information technology has contributed to increasing polarization in the labor market: IT displaces many middle-class jobs, like those in the car industry, that require moderate levels of training and skill, and the new jobs are either high-skill high-wage or low-skill low-wage jobs, contributing to the hollowing out of the middle class that is so frequent a topic of discussion.

Second, in the United States, the share of GDP going to workers has been steadily declining. Third, the increasing concentration of “super-star” firms in certain sectors means that they can erect barriers to entry that reduce competition, in ways that limit the bargaining power of both workers and customers (Zago, 2019, citing Autor et al., 2017). And, fourth, one factor in the creation of the modern precariat is the fact that people within the United States are less likely than you might have expected to go to where the jobs are: and one cause here is the increasing

disparity between the costs of living in rural areas and small towns, on the one hand, and the most productive metropolises, on the other . . . just as another is the polarization of values between the small town and the cosmopolitan city.

The policies that have been considered to meet these difficulties aim, in effect, to strengthen the position of workers, in one of four ways. First, by increasing their skills, through education and training. Second, by assisting them in identifying new opportunities. Third, through a commitment by the government to be an employer of last resort, guaranteeing people a meaningful job consistent with their developed capacities. And fourth, by guaranteeing a basic income, which allows people to refuse jobs that are not sufficiently rewarding in income, esteem, sociability, or significance.

The first three of these possibilities treat the problem as a matter of reforming the nature of work. But the last entertains the possibility of sharing the social product in ways that move beyond the idea of work as the temporal and eudemonic center of our lives. This, then, is the post-work option; and one possible such option is to guarantee everyone a basic income.

6.

The sociologist David Frayne has taught me about many problems with and at work in the North Atlantic world explored in the sociological literature. Let me add just a few to those I have already mentioned: the current labor market leaves many people without incomes adequate for a decent life; life without work is stigmatized, “overshadowing the value of noninstrumental activities like care, leisure, play, or learning for its own sake”; too many important social contributions are not recognized at all, “particularly care work and domestic labor, which are unrecognized, unremunerated and unequally distributed” (Frayne, 2019). If work is serving us so many of us so poorly, it is a natural thought that we might take the great wealth produced by our automated society and use something other than the labor market to share it. That’s one reason we have

seen increasing support for the proposals for a universal basic income that I mentioned just now.

It's worth pointing out one key feature of a world in which the distribution of the social product is not done only by the labor market. Wages and the other benefits of work are differential today in large measure because they channel people into tasks whose products (whether goods or services) meet a demand that is measured by a price. As any economist will tell you, the result is that wages, like prices, integrate and reflect a great deal of information both about what people want to have and about what they want to do. A world in which everyone had an equal reliable basic income would be a world in which no one would have to work to meet their basic needs. Like current work, such employment as there was in such a world would have to be incentivized: the incentives would include, as they do for decent jobs today, the sociability and significance of the work world as well as financial rewards. We could reward the socially necessary tasks that cannot be automated and that people do not find intrinsically rewarding with large hourly wages and, especially if we combined them with shorter hours, a significant part of the population could increase their incomes with a few hours of this work, while bearing a smaller share of the costs in unpleasantness, which, in today's economy, are concentrated in the working lives of a few. But many people would be making additional income in other ways—as artists, say, but also by selling things such as wool and vegetables like Marx's part-time farmers. A universal basic income would reshape the economy, but by definition it wouldn't produce a world without money. And so it would continue to produce the unequal wealth and income that are bound to be a feature of any market economy. What the results of all this would be is hard, I think, to imagine: and since the universal basic income would have to be funded somehow, the actual effects would depend very much on things like the progressivity of the income tax system, and whether or not there were taxes on capital and on inheritances, and so on. So, normative questions about income inequality would remain, even if one of the challenges of the present world—the fact that too many people do not have enough for a dignified human life—were solved.

7.

There may be normative reasons, as the philosopher Denise Celentano has taught me, for wondering whether people might not have a duty to work if they can.⁷ This is not just a matter of adhering to the secularized version of the Protestant Ethic that pervades many modern societies. The most natural understanding of the feature of “work” that is relevant here is that—at least if the work is not a bullshit job—it entails spending time doing something that makes a social contribution. Once we think, as Rawls taught us to do, of the basic structure as a scheme of cooperation, whose obligations and rewards need to be fairly distributed, then someone who is not making a contribution to the scheme is no more entitled to its privileges than someone in a society elsewhere. It does not follow, of course, that we owe her nothing; any more than the fact that someone lives in another country means we can ignore her in our moral thinking. But there seems to be a basis here for the thought that it is only through work that we are connected to others in society in the ways that raise the question of distributiveness fairness at all.

There are immediate reasons for resisting this argument for a duty to engage in paid work, though. For one thing, the basic structure includes more than the economy. The family is a site where many of the things we do, especially in the domain of child-rearing, have not, at least in the past, counted as work. And so is our political life as participants in public reasoning, as voters, and through our respect for the laws. Here, too, we contribute in ways for which we have not historically been rewarded through income; though in politics, as in the family, we can earn esteem for our contributions. It is actually quite hard to imagine a modern person whose life is totally without contributions in one or other of these domains away from work. But even if someone succeeded in escaping from contributing, there would be other reasons for wondering whether a duty to work—imposed as a condition of any social provision—would be consistent with other things that matter.

It seems evident, for example, that all of us have obligations to others—most obviously, negative duties to avoid unnecessary harm—that

⁷ My discussion here is especially dependent on the work of Denise Celentano.

do not depend on the fact of our being connected in a scheme of cooperation. Furthermore, at least some ways of making social provision dependent on contributions may violate ethical notions of autonomy. Beatte Rössler writes:

Subjects have to work, whether they want to or not. With regard to the very question of why people work their autonomy does not seem to play a role. (Rössler, 2012, p. 77)

She's asking us to consider whether imposing work as a condition of providing someone with the means for a dignified existence might violate their autonomy. But, as Joseph Raz has argued, autonomy needs only "an adequate range of options" and to "be independent from coercion and manipulation by others" (Raz, 1988, pp. 389–90). Why should having to do some kind of work or other (picked, let us suppose, from a wide range of meaningful occupations for which you are prepared) mean you don't have adequate options and freedom from manipulation? To say that someone who is "forced" in this way to work is "coerced" or "exploited" is to beg the question. For to be coerced is to be forced in morally impermissible ways to do something, and to be exploited is to have someone wrongfully take advantage of your vulnerability. A person who has the option of many decent jobs, even if she'd rather do nothing, is not vulnerable to any particular employer, nor, given a range of reasonable choices, is it evidently wrong to expect her to accept one of them. To put it simply, it's not obvious that to have an adequate range of choices you must have the option of relying on the labor of others for your basic needs. My point, here, is not to decide who is right, but to insist that there is scope for philosophical argument about this issue.

There are reasons, too, for wondering about the social and psychological challenges of a world in which many fewer people are gaining their incomes from work. The sociological literature on employment raises many doubts about the possibility of a satisfying life without it. But, on the other hand, evidence drawn from the experience of unemployment—or of retirement—in our current social system seems a bad place to start. As Daniel Sage (2018, p. 207) has argued recently, "unemployed people live in societies where paid work yields status,

identity, respect and human worth.” “The damage of unemployment,” he argues, “is thus not the absence of paid work but the failure to conform to a powerful social norm.” In a culture, where, as the political scientist James Chamberlain argues, hard work is seen as “an expression of virtue and good character, . . . symbolizes independence and is the main way to fulfill civic duty and make a social contribution,” it is not surprising if worklessness produces depression or anxiety (Chamberlain, 2018, p. vi).

Paul Gomberg has developed an account of what he calls “contributive justice,” which is based on the thought that the proper interpretation of egalitarianism requires us to think about equality not in what we get from but in what we give to social arrangements. His basic argument is simple:

Income and wealth are distributed either unequally or equally. If unequally, then those with less are unjustly subject to social contempt. But equal distribution is impossible because it is inconsistent with bargaining to advance our own good. Hence justice in distribution of income and wealth is impossible. (Gomberg, 2016, p. 31)

Elsewhere he writes that “philosophers have thought that justice is about what people get; I think it is about what people are able to do, particularly how they are able to develop their abilities, give back to society, and be respected for their contributions.” So, we should “share labor, including the boring work most of us like to avoid if everyone is to have an opportunity to develop all of their abilities” (Gomberg, 2007, p. vi). But I confess it is simply not clear to me why reasonable financial inequalities entail contempt for the worse-off. Indeed, I have argued to the contrary in *The Lies That Bind* (Appiah, 2018, p. 135 *et seq.*). Once more, though, my aim is to identify a philosophical question, not to settle it.

8.

In a well-known essay on “Time, Work Discipline, and Industrial Capitalism,” E. P. Thompson explored the way that industrialization created new patterns in the use of time, inculcating in the working classes the idea (which Weber had seen at the Protestant root of modern

capitalism) that “time is money.”⁸ And he suggested, toward the end, that if what he called “the purposive notation of time-use” became “less compulsive, then men might have to re-learn some of the arts of living lost in the Industrial Revolution: how to fill the interstices of their days with enriched, more leisurely, personal and social relations; how to break down once more the barriers between work and life” (Thompson, 1967, p. 95).

If we are to make these changes, it seems evident that we will need an education system that readies all of us for a life that is not structured by the necessity of labor. Our current models of a liberal college education, as I mentioned earlier, give thought to preparation not just for work or even for citizenship but for the whole of life. But that form of education is focused on the part of the population that had the skills and the will to engage in the sort of learning that our colleges and universities currently offer. And, of course, since they assume that we will, in fact, be workers, they do not focus on preparing us for a life in which, for at least much of the time, we are engaged in tasks we perform for no instrumental rewards.

All these issues will arise even if a post-work society is just a less-work society, in which we do paid work for many fewer hours; *or* if we settle on providing a universal basic income and many people engage in no paid work at all; *or* if we share rewarding labor or the unpleasant socially necessary tasks that cannot yet be automated or both. Work will need to be refigured if we take seriously the idea of lives in which income is no longer dependent on a job. At the heart of these reflections is the recognition that, just as the Industrial Revolution produced new conceptions of value (like the equation of time with money), so in our modern economy, changing institutions will have to be accompanied by conceptual and institutional innovations that it will take imagination to shape and to share.⁹

⁸ Max Weber quotes a passage from Benjamin Franklin’s “Advice to a Young Tradesman, Written by an Old One,” early in *The Protestant Ethic*: “Remember that TIME is Money. He that can earn Ten Shillings a Day by his Labour, and goes abroad, or sits idle one half of that Day, tho’ he spends but Sixpence during his Diversion or Idleness, ought not to reckon That the only Expence; he has really spent or rather thrown away Five Shillings beside” (Weber, 2001, p. 14).

⁹ I am grateful to the Berggruen Foundation for the funding for three postdoctoral fellows—Denise Celentano, David Frayne, and Ricardo Zago (a philosopher, a sociologist, and an economist)—who spent the academic year 2018–19 with me at NYU to begin thinking about these questions.

Bibliography

- Acemoglu, D. & Restrepo, P. (2017). "Robots and jobs: Evidence from US labor markets." NBER Working Paper No. 23285 issued in March 2017.
- Anderson, E. (2017). *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*. Princeton, NJ: Princeton University Press.
- Appiah, K. A. (2017). *As If: Idealization and Ideals*. Cambridge, MA: Harvard University Press.
- Appiah, K. A. (2018). *The Lies That Bind: Rethinking Identity*. New York: Liveright.
- Arendt, H. (1998). *The Human Condition*. Chicago, IL: University of Chicago Press.
- Autor, D. H., Dorn, D., & Hanson, G. H. (2013). "The China syndrome: Local labor market effects of import competition in the United States." *The American Economic Review*, Volume 103, Number 23, 2121–68.
- Autor, David, Dorn, David, Katz, Lawrence F., Patterson, Christina, & Reenen, John Van (2017). "The fall of the labor share and the rise of superstar firms." <https://economics.mit.edu/files/12979>.
- Bessen, J. E. (2016). "How computer automation affects occupations: Technology, jobs, and skills." <https://www.bu.edu/law/working-papers/how-computer-automation-affects-occupations-technology-jobs-and-skills>.
- Celentano, D. (2019). "The normative problem of work: An overview." *ms*.
- Chamberlain, J. (2018). *Undoing Work, Rethinking Community: A Critique of the Social Function of Work*. Ithaca, NY: Cornell University Press.
- Frayne, D. (2019). "The hard problem." *ms*.
- Goldin, C. & Katz, L. F. (2018). "The race between education and technology." In: *Inequality in the 21st Century*, Grusky D. & Hill J. (eds.). New York: Routledge, pp. 49–54.
- Gomberg, P. (2007). *How to Make Opportunity Equal: Race and Contributive Justice*. Alden, MA: Blackwell Publishing.
- Gomberg, P. (2016). "Why distributive justice is impossible but contributive justice would work." *Science & Society*, Volume 80, Special Issue: Socializing Philosophy, 31–55.

- Graeber, D. (2018a). *Bullshit Jobs: A Theory*. New York: Simon and Schuster.
- Graeber, D. (2018b). "Bullshit jobs: Why they exist and why you might have one." Sean Illing interview with David Graeber for *vox.com*. October 10. <https://www.vox.com/2018/5/8/17308744/bullshit-jobs-book-david-graeber-occupy-wall-street-karl-marx>.
- Graeber, D. (2018c). "Nearly half of you reading this have bullshit jobs." Nick Romeo interview with David Graeber in *The Daily Beast*, May 18. <https://www.thedailybeast.com/nearly-half-of-you-reading-this-have-bullshit-jobs>.
- Hacking, I. (2002). *Historical Ontology*. Cambridge, MA: Harvard University Press.
- Honneth, A. (1996). *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Cambridge, MA: MIT Press.
- McIntosh, D. (2011). "How to tell summer is here: It's union picnic season." *newlaborpress.org*, June 1. <https://nwlaborpress.org/2011/07/picnics>.
- Patterson, O. (1992). *Freedom in the Making of Western Culture*. New York: Basic Books.
- Rawls, J. (2001). *Justice as Fairness: A Restatement*. Erin Kelley ed. Cambridge, MA: Belknap Press.
- Raz, T. (1988). *The Morality of Freedom*. Oxford: Oxford University Press.
- Rössler, B. (2007). "Work, recognition, emancipation." In *Recognition and Power: Axel Honneth and the Tradition of Critical Social Theory*, Van den Brink B. and Owen D. (eds.). Cambridge: Cambridge University Press, pp. 135–164.
- Rössler, B. (2012). "Meaningful work: Arguments from autonomy." *Journal of Political Philosophy*, Volume 20, Number 1, 71–93.
- Sage, D. (2018). "Unemployment, wellbeing and the power of the work ethic: Implications for social policy." *Critical Social Policy*, Volume 39, Number 2, 205–28.
- Schwab, K. (2016). "The Fourth Industrial Revolution: What it means? How to respond?" *Global Agenda*, World Economic Forum, January 14. <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond>.
- Thompson, E. P. (1966). *The Making of the English Working Class*. New York: Vintage.

Thompson, E. P. (1967). "Time, work-discipline, and industrial capitalism," *Past & Present*, Volume 38, 56–97.

Weber, M. (2001). *The Protestant Ethic and the Spirit of Capitalism*, Stephen Kalberg (trans.). Chicago, IL: Fitzroy Dearborn Publishers.

Zago, R. (2019). "On the future of work." *ms*.

2

Political Epistemology and Social Critique

Sally Haslanger

1. Introduction

My recent work defends a practice-based account of ideology. Social practices and social structures depend on a collection of social meanings—I call this a *cultural technē*—that provides a “stage-setting” for action and is a constituent part of the local social-regulation system (e.g., Haslanger 2018, 2017a, 2017b). These cultural technēs enable us to coordinate by providing the paths and signals of our practices. An ideology is a cultural technē “gone wrong.” It prevents us from recognizing or creating forms of value and/or organizes us in unjust ways. This account of ideology is functionalist, pejorative, but not doxastic. It is functionalist because the evaluation of a technē as ideological depends on how it functions in a context; it is pejorative because, in being ideological, it functions to create or sustain injustice; and it is not doxastic because a cultural technē is not a set of beliefs, but is, rather, a set of public meanings (though some parts of it may be internalized as beliefs and other attitudes).¹ The cultural technē both provides resources to interpret and shapes the material world.

Throughout this work I have relied on the idea that some (but not all) cultural technēs are ideological. My task here is to say more about the

¹ I am an externalist about linguistic meaning (meanings aren’t in the head), and although a model of social meanings of the sort I have in mind would have to be more complicated than standard models of linguistic meaning, some of the basic tools from externalist philosophy of language can be applied more broadly. I also consider signs and symbols to be part of the technē (just as letters and words are part of our linguistic technē). See also (Haslanger 2020a).

basis for such normative evaluation of social meanings and the practices they enable. Ideology critique is an important part of efforts to promote social justice, but how is critique possible and how is it warranted?²

In what follows, I will sketch three problems for ideology critique, drawing on the work of Robin Celikates (2016): the normative challenge, the epistemological challenge, and the explanatory challenge. Before attempting to address these challenges, I will situate the inquiry as a form of non-ideal social theory. My project, however, is not to define or develop non-ideal theory generally, but to sketch a form of critical theory that has emerged in the context of social justice movements—broad movements that include participation from activists, academics, artists, and ordinary folk attempting to live their lives with integrity and hope.³ Members of such movements engage in critique as agents *within* a set of unjust social practices. Critique of a practice is aimed at others who are engaged in the practice with us; the question is how we should go on together from here.

I argue that under conditions of ideology a standard model of normative political epistemology—using a domain-specific reflective equilibrium—is insufficient. (See also Haslanger Forthcoming.) Moreover, simply including diverse knowers as sources of situated knowledge, taken at face value, is also insufficient. A *critical* standpoint is necessary. One way of achieving a critical standpoint is through consciousness raising (CR); consciousness raising offers a paradigm shift that cannot simply be derived from our existing moral framework. However, not all epistemic practices that appear to “raise” consciousness are warranted.⁴ Nevertheless, under

² In this text, I will speak, generally, of “social justice.” However, it is reasonable to question whether ‘justice’ is the right term for what emancipatory movements seek. (Thanks to Lorna Finlayson for raising this issue.) As I understand the phrase ‘social justice,’ it does not presuppose that what’s at stake is captured, or even aimed at, by contemporary theories of justice. I will not enter explicitly into this debate here, and will leave the notion open-ended.

³ The term ‘critical theory’ has multiple uses, and sometimes functions as a proper name for a particular method, e.g., Frankfurt School Critical Theory, or Critical Race Theory. I use the term ‘critical theory’ (lower case) for an activity within social justice movements that resist oppression on the basis of class, race, ethnicity, gender, sex, sexuality, nationality or immigrant status, disability, size, etc. I am not attempting to regiment or define a method, but to illuminate aspects of an approach that I believe has been neglected in mainstream political philosophy.

⁴ One might use the term ‘consciousness raising’ as a success term—one’s consciousness isn’t raised unless the result is a warranted critique—or simply as a fallible process or procedure. Sarachild (1978) gives an example of the latter: “A recent *New York Times* article referred to a meeting called by Henry Kissinger to talk to the executives of the major television networks

certain conditions that I aim to specify, consciousness raising produces a warranted critical standpoint and a *pro tanto* claim against others through a process of inquiry.⁵ Even so, the political work remains: under conditions of collective self-governance, there is no guarantee that all warranted claims can be met simultaneously. There will be winners and losers even after legitimate democratic processes have been followed (Allen 2004). A warranted critique of society may not be politically successful; there are no guarantees. The political basis for conflict resolution between *pro tanto* claims is not my topic here.

2. Ideology and Subjection

To motivate the problem, it is important to say more about the epistemic impact of ideology.⁶ Althusser (1971) distinguishes *repressive state apparatuses* (RSAs) and *ideological state apparatuses* (ISAs). RSAs include the “government, administration, army, courts, prisons” that “function by violence” or “massively and predominantly by repression.” ISAs include religion, education, the family, the legal system, the political system, trade unions, communications/media, and culture (“literature, the arts, sports, etc.”) that “function massively and predominantly by

about the content of their programs as a ‘curious “consciousness-raising” session’ with a Secretary of State” (147). (Note: the article (Brown 1974) actually says the meeting “was described by one who attended as a ‘consciousness-raising session’ on certain world issues that Mr. Kissinger believed deserved studious attention.”) If one opts for using it as a success term, then some practices that look a lot like consciousness raising are not genuine cases because consciousness isn’t actually “raised.” Sarachild (1978) and others insist that what defines CR is not a precise procedure, but results (147). As will become clearer as I proceed, I will use the term ‘consciousness raising’ to refer to a variety of practices that give rise to a paradigm shift in understanding one’s social circumstances; the new paradigm provides participants an “oppositional consciousness.” However, not all paradigm shifts are warranted and not all forms of oppositional consciousness provide us insight into justice. I will follow the feminist tradition, however, in assuming that consciousness raising is successful only if it yields a warranted critique.

⁵ There is a substantial literature in epistemology on the relationship between warrant, justification, and entitlement and there are substantive differences in how the terms are used. I am not going to delve into that discussion in this chapter. I will mostly use the term ‘warrant’ rather than ‘justification’ because I want to distance myself from the internalist and doxastic assumptions that tend to be associated with justification (Pollock and Cruz 1999). I do not, however, have a theory of warrant.

⁶ This section draws on (Haslanger 2019a), which also expands some of the points I make here.

ideology.” (No state apparatus is purely one or the other, and each depends crucially on the other, though in modern society, the ISAs are the dominant mode of social management.) A crucial difference between an ISA and an RSA is that individuals are hailed into a subject position by an ISA (Althusser calls this a process of “interpellation”), rather than violently forced into it. It is characteristic of those “good subjects” who respond to the hailing that they take up the norms as binding on themselves, so they don’t need to be coercively managed. For example, to maintain a division of labor, instilling literacy, numeracy, and other kinds of technical “know how” is not sufficient:

besides these techniques and knowledges, and in learning them, children at school also learn the ‘rules’ of good behaviour, i.e. the attitude that should be observed by every agent in the division of labour, according to the job he is ‘destined’ for: rules of morality, civic and professional conscience, which actually means rules of respect for the socio-technical division of labour and ultimately the rules of the order established by class domination. They also learn to ‘speak proper French’, to ‘handle’ the workers correctly, i.e. actually (for the future capitalists and their servants) to ‘order them about’ properly, i.e. (ideally) to ‘speak to them’ in the right way, etc.

(Althusser 2014/1971, 235–236)

The local ISAs interpellate subjects so that they perform the practices of their social milieu freely: this is how things are; this is what we do; this is who we are. As Althusser emphasizes, the good subjects “work all by themselves”! This conception of ideology (though not in these terms) is also a theme in Foucault’s work (e.g., 1979, esp. ch. 5, and for a feminist application see (Bartky 1990)).

My conception of ideology is Althusserian. We participate in social practices guided by a set of public meanings, scripts, etc. Particular practices are signaled and structured by features of the material conditions. A blackboard and desk or podium marks the front of a classroom. We organize ourselves in such a space depending on our role in that setting. The front of the classroom has a meaning that both students and teachers understand, and guides them in the activity of learning together.

The network of meanings that play a role in interconnected practices form a *cultural technē*.⁷ The boundaries of a cultural technē are not precise. The cultural technē of a philosophy classroom tends to be quite different from a gender studies classroom; the cultural technē of Harvard is quite different from the cultural technē of MIT. Those who are socially fluent in a particular setting have internalized its cultural technē—they are its “good subjects.”

We are “hailed” into practices in a variety of ways, e.g., we are hailed into speaking English by having English spoken to us; we are hailed into the role of student by being sent to school and finding ourselves responding to the teacher as an authority (nudged by coercion); we are hailed into adulthood by having to pay the rent (with threat of coercion in the background). We then develop ways of being and thinking so that we are (more or less) fluent English speakers, fluent students, fluent rent-paying adults. Ideology is not a set of beliefs, though it may produce belief about what is apt or inapt, right or wrong, and related desires, emotions, and other attitudes. As Althusser says, “Ideology always exists in an apparatus and its practice or practices. Its existence is material” (Althusser 1971, 259). The world around us is structured so that we typically embody a practice before we even know we are engaged in it (McGeer 2007; Zawidzki 2013).

Social practices organize us around things taken to have more or less value; let’s call these (assumed or constructed) *sources* of value and disvalue.⁸ Some sources are material (such as medicine, traffic, toxic waste), and others not (such as time, knowledge, boredom). For example, the practice of attending an academic lecture organizes us around a presumptive source of knowledge. The cultural technē of academia,

⁷ In the past I have used the term ‘schemas’ both for public *cultural* schemas and internalization of them as *psychological* schemas. This has caused confusion, so I now use the term ‘social meanings,’ and for webs of meanings, ‘cultural technē.’ ‘Social meanings’ include narratives, patterns of inference, default assumptions, symbols, and other cultural memes that one might not normally consider “meanings” in a narrow sense. See also (Haslanger 2018).

⁸ Following Giddens and Sewell, I originally employed the term ‘resources’ in this context. The term ‘resource,’ however, has a positive connotation and I’ve been urged to find another way of speaking of resources that more easily includes things taken to have negative value. (Thanks for this nudge to Jeffrey Stout.) Until I find something better, I will use ‘sources’ with the understanding that sources come in many different forms. Note that because we are not assuming that what we “take to be” of value or disvalue is correctly valued, we should not assume that a ‘source’ *actually* has the value or disvalue attributed to it.

overall, has value, though some parts of it, or its manifestation in some settings, may be ideological. A cultural technē can go wrong in different ways. It may distort our capacity to value, i.e., to recognize what is truly valuable and what not (Anderson 1993, ch. 1); it may organize us in response to presumed value in unjust ways. For example, recent work on epistemic injustice argues that academic practices place unwarranted restrictions on who counts as a knower, what form knowledge must take, and the legitimate sources of knowledge. (See Tuana 2017 for a useful overview.)

Under conditions of ideology there is, by hypothesis, a range of unjust social practices that oppress a group; however, not everyone experiences the oppression as such. Those who are fluent in the practices may not even recognize them as social practices, e.g., a practice may be naturalized or taken for granted. The working class may not recognize their exploitation as such; women may not agree on what practices are sexist. And even a problematic practice may be experienced as valuable and produce something of value. At the very least, practices enable coordination; coordination, even on non-optimal terms, is valuable because coordination on any terms is important and can be difficult to achieve.

The epistemic position of the “good subject” is complicated. As just mentioned, some subjects embedded in unjust practices do not experience them as unjust; others may have a vague dissatisfaction; and others may have an articulated critique.⁹ Even being deeply critical of a practice does not prevent one from being fluent in it, and because resistance is often punished, many will have reason to comply with practices they abhor. Moreover, there may be no better live option. Because we depend on coordination with others, we are often just stuck with enacting an unjust system we are embedded in, for lack of better alternatives.

The term ‘standpoint epistemology’ emerged as an effort to address the problem of ideology: Where does one stand to critique ideology?

⁹ There is an important set of questions about the kind of knowledge gained by those who occupy a subordinated position and its relationship to critique. Patricia Hill Collins argues that “Black Feminist thought rearticulates a consciousness that already exists [among Black women]” (1989, 750). I am not denying that often the resources for critique already exist in the experience of the subordinate; my claim is that sometimes, for some groups or in relation to some practices, complicity in unjust practices is deeper than this suggests. See also Khader (2011).

Because ideology, when successful, recruits us into fluent participation in an unjust structure, some of those who are subject to subordination will not develop or accept a critique of it. This has two important consequences. First, “situated knowledge”—knowledge gained by virtue of occupying a particular social position—need not itself be sufficient to generate critique. So, some broadly empiricist claims that “all knowledge is situated,” although true, will not necessarily be enough to develop a critical perspective; an effort to include diverse knowers in inquiry, although important, will not be enough to disrupt ideological practices. (See also Intemann 2010.) Situated knowledge may just provide knowledge of the practice, without knowledge of what makes it problematic or what would be better. Second, there will not be an easy consensus among the subordinated to challenge the status quo. Further, critique may take aim at deeply held identities that both enable one to coordinate with others and provide a basis for self-esteem.

How does a critic proceed? In principle, if we know what is just and unjust, then the proper target of ideology critique simply follows: we should disrupt the cultural *technē* that prevents us from valuing things aptly and disrupt those social structures that produce injustice. However, because knowledge is situated, and so distributed (I assume that this includes knowledge of value and other sorts of normative knowledge), individual acts of reflection, or reflection by a group of similarly situated knowers, is not enough. And because ideology masks injustice and blocks us from an understanding of our social milieu, how can we develop a warranted critique? Robin Celikates (2016, 3–4) elaborates three challenges an account of ideology critique must address.

- (i) *Normative challenge*: What makes an ideology problematic? Are there objective moral truths by reference to which we can judge a social arrangement defective or unjust? If so, how do we gain knowledge of those truths? If not, then on what basis do we undertake critique?
- (ii) *Methodological or epistemological challenge*: From what standpoint does the critic speak? Traditionally critical theory is embedded in a social movement and aims to articulate the interests and demands of the oppressed. But then the question is “which

insights of which agents—given that they usually do not constitute a homogeneous category—the critical theorist articulates.”

- (iii) *Explanatory challenge*: If an ideology functions at the level of a system, “what exactly holds the rather broad conglomeration of partly psychological, partly social mechanisms—from implicit biases via stereotypes to looping effects—together and makes them into elements of one ideology”? And how does critique disrupt the systematic injustice sustained by the ideology?

To address these challenges, it would appear that we must provide a full-blown social theory, moral metaphysics, and moral epistemology. Fortunately, there is a narrower task that we can begin with.

3. Methodological Preliminaries: Narrowing the Task

a. Moral Truths

We are not starting the normative inquiry from scratch. Those engaged in justified political resistance cannot avoid the claim that there are *some* moral truths. So it is not my task to argue for an objective basis for moral judgment.¹⁰ Moreover, ideology critique *is critical*; it does not make a claim about the nature of justice. What counts as ideology is a matter of the injustice of its effects and the (bad) values it promotes/embodyes, so it focuses on identifying injustice and harm.

This move to focus on injustice is a strategy common to many forms of non-ideal theory. The form of critical theory I am developing here, however, goes further than simply a *focus* on injustice; it resists the call to provide a theory of justice by reference to which we can judge cases to be just or unjust. We do not need to *know what justice is* or have a complete moral theory to engage in social critique. We can know, in some cases, when a practice is unjust, without knowing why it is; in fact, the standard form of normative theorizing depends on the validity of such pre-theoretical judgments, e.g., we do not need a moral theory to tell us

¹⁰ I leave open the meta-ethical view about the nature of moral facts.

that slavery or rape is wrong and standard justifications of our moral theories depend on the adequacy of such judgments. Moreover, modal knowledge of what makes something just or unjust (which presumably is what a theory of the nature of justice provides) is not required to remedy instances of it. And finally, injustice may not be a proper kind, so attempting to construct a theory of justice may lead us astray, causing us to neglect forms of injustice that don't fit our theory (Young 1990, ch. 2).

The resistance to articulating an "ideal theory" of justice is methodologically deeper than this, however.¹¹ In the case of some kinds worthy of our attention, there are pre-existing and projectible regularities that we have reason to identify and investigate (water is H₂O; the highest poverty rate in the US by race occurs among Native Americans (in 2018 = 25.4%)).¹² But in the social world, the adequacy of our conceptual framework should not simply be judged by the facts it captures, but by *what it does*: the resources it provides for organizing and understanding ourselves. For example, how should we define *refugee*? Our chosen definition matters for people's lives. In the social domain, the direction of fit goes both ways: we aim to capture facts about the world, and our doing so can contribute to producing facts—sometimes facts we are trying to capture and sometimes new ones (Hacking 2002; Haslanger 2012). The social effects of a proposed definition (or theory, more broadly) is a consideration that tells for or against it; there is a sense in which a definition can be unjust.

There are situations when we can simply stipulate a new word to do the conceptual or linguistic work we need done: 'super-spreader' or 'coronnials.'¹³ But in philosophically interesting cases, there is something we are aiming to understand that is not simply constituted by what we decide—by stipulation of meaning—yet at the same time is not completely free of the discursive tradition that provides the tools to identify it. We are seeking an understanding of practices in which we are currently engaged as participants. The practices are not fully understood,

¹¹ I discuss this also in (Haslanger 2019b) in relation to understanding what race is and in (Haslanger 2020c) in considering conceptual engineering more generally. See also (Moi 2015; Mills 2005).

¹² <https://www.povertyusa.org/facts>.

¹³ <https://www.latimes.com/world-nation/story/2020-04-11/coronavirus-covid19-pandemic-changes-how-we-talk>.

however; and they are open-ended, revisable, possibly self-defeating. In making sense of them, we are making judgments about how to better understand what we are doing, and how then to go on; this affects what the practice is. This process involves an evaluation and extension of claims we make about, and within, the practice and the vocabulary we use to do so. However, it is not primarily a linguistic exercise: we aren't just deciding how to use existing terminology, but how to collectively orient ourselves toward the world and each other. In other words, we are situated inquirers, and the question is how we should go on from here (see also Walker 2007). Because of the influence of ideology, in particular, we should be cautious about dominant ideas about where we have been, where we are now, and how or whether we should continue. There may be some practices, e.g., medicine or mathematics, in which we can rely mainly on experts to decide the best way to go on. But in the social domain, we should figure this out collectively.

b. Anti-Utopianism

Resistance to ideal theory is also supported by the fact that there are many ways to organize social life, so the goal is not to ask what practices are the *best* way. For example, the production and distribution of food is part of social life, and there are better and worse ways to do it. It would be misguided to argue that three square meals a day is the best way to organize food consumption and all others should be judged by reference to this ideal; but some ways of producing and distributing food are unhealthy, some exploit workers, some destroy ecosystems, some cause suffering and/or death of sentient beings. We need not assume that there is a *best* way to organize food production, consumption, and disposal in order to engage in critique. A normative social theory provides tools to identify ways in which our current practices are inadequate so we can do better. What counts as better will depend hugely on local factors, e.g., the geography, economy, cultural traditions, and human biology.¹⁴ In

¹⁴ This is why relying on the state to manage social practices is often counter-productive. See Ostrom (1990) and Agarwal (1992, 2009).

existing societies, injustice is already rampant. Rectification is a priority. For this reason, a normative social theory is *anti-utopian* (though it does require imagination and a kind of hope (Solnit 2016)).

Moreover, because the focus is on the inadequacy of current practices in a particular context, we must be sensitive to the fact that practices occur within cultural traditions that are a source of value. This is not just to say that people value their ways of life. Practices are valuable and produce goods that are internal to the practice. For example, in a division of labor, there will be a division of expertise. Even if the division of labor is unjust, individuals marked for a particular set of tasks may develop a set of skills, a sensibility, solidarity with each other, by virtue of participation in a particular position in the structured set of practices. Goods may accrue to the participants, even if overall the structure is problematic and should be changed. In such cases, sacrifice is inevitable.

One may be concerned that once we recognize that there are many acceptable ways to organize social life and that our critique is always situated, we must give up on the objectivity of value. However, objective values need not be ahistorical or acontextual; they may be path-dependent. What's valuable depends, *inter alia*, on what is available to value. Jack Balkin makes this point:

Values are not so much what humans have as what they do and feel. Human beings possess an inexhaustible drive to evaluate, to pronounce what is good and bad, beautiful and ugly, advantageous and disadvantageous. Without culture, human values are inchoate and indeterminate; through culture they become differentiated, articulated, and refined. (1998, 27–28)

In developing this point, Balkin relies on examples of aesthetic value: the creation of different sorts of musical instruments and the different configurations of sound they produce enables us to cultivate different ways of hearing and attaching aesthetic value to music (1998, 28). It would not be possible aesthetically appreciate rock and roll without the invention of electric guitars; our aesthetic sensibility evolves in response to the creation of new sounds. Balkin extends this to moral sensibility and moral value:

We concretize our indeterminate value of justice by creating human institutions and practices that attempt to enforce it and exemplify it, even (and especially) if we recognize that all of these institutions are imperfectly just. Of course, because justice is an indeterminate standard, there is no necessary way to exemplify it. The value of justice does not tell us, for example, whether a democratic legislature should have one, two, or three houses. Hence the institutions that people construct to exemplify justice may be different in different eras and different lands. (1998, 30)

I take it that Balkin's point is not just that there may be different ways of exemplifying well-defined abstract rules or procedures (be they aesthetic or political), but that our ultimate values are indeterminate and any attempt to render them determinate will be specific to cultural and material conditions, so may not be transposable to other conditions. (See also Khader 2019.) It is compatible with this, however, that given the socio-historical conditions, there are objective truths about what is beautiful or just in the context in question.

The considerations just sketched suggest two dimensions of indeterminacy, one diachronic and the other synchronic, that pull against the impulse to produce an ideal theory of justice. Synchronically, any "sense of justice" or appreciation of the "value of reciprocity" (etc.) that might provide a basis for moral theorizing across social and cultural differences is indeterminate (Kymlicka 2002, 2–4; Dworkin 1977, 179–183); and our efforts to articulate it in a way that renders it determinate will inevitably incorporate particular socio-historical elements that make it apt for some contexts, or some communities, but not others.¹⁵ Diachronically, even if

¹⁵ I have argued elsewhere (Haslanger 2020c) that the standard model of domain-specific reflective equilibrium of the sort recommended by Rawls and Scanlon suffers from a kind of status-quo bias that makes it unsuitable for ideology critique. I grant that reflective equilibrium of the broadest sort—including all of our beliefs (empirical, logical, metaphysical, normative, etc.)—is the best we can aim for in inquiry, but I reject the idea that moral inquiry should be pursued by aiming for a narrower reflective equilibrium between normative judgments and principles and "uncontroversial" non-normative claims (Scanlon 2003). I also resist the idea that moral theory should be treated as if it is undertaken by an individual deliberating about what is right or what to do (which is what Scanlon (2003) recommends). Moral inquiry, like scientific inquiry, is a collective project and conclusions that may be warranted for an individual may not be warranted as part of the collective effort to decide how we should live together.

we are able to specify a conception of justice that is fully determinate and applies generally (for current purposes), we are not in a position to grasp the full range of possibilities that we might face and decide in advance what would be appropriate under radically different conditions. The world poses challenges that our previous (or existing) understandings and sensibilities do not solve. Who could have imagined, even a century ago, the morally significant possibilities created by the biological sciences (assisted reproductive technology, cloning) and engineering (automobiles, space travel, cellphones, robotics)?¹⁶ Our sensibilities evolve in response to new conditions, and the evolving sensibilities—and critique of those sensibilities—is part of a process of determining what is just here and now. It may be that we should attempt to develop theories that are apt for our current conditions based on our current knowledge. But the point I am making is not just that we are fallible, i.e., that there is a truth about the nature of justice (for all times, all conditions) out there waiting to be found and we have only fallible access to it; the claim is that what is just or unjust does not float free of our sensibilities and our practices, and the relevant practices are, like other practices, open-ended and revisable (in the aims, procedures, and results), and depend on our collective and critical efforts to go on, together, from here.

A cultural *technē* is an evolving specification of our “inchoate and indeterminate” drive to evaluate in response to our material (biological, geographical, economic) conditions. To suppose that we can articulate an ideal that is not conditioned by our cultural *technē* and, even if we could, that it could speak to us is implausible. This does not leave critique without normative resources. Social critique can, at the very least, draw on our inchoate and indeterminate sense of justice and its articulation in other contexts to construct and demand a better alternative to the current practices. The fragmentation of our social practices and relative (but incomplete) autonomy of social systems generate tensions and contradictions that can prompt reflection and reconfiguration of our normative resources.¹⁷

¹⁶ For an excellent discussion of the impact of the invention of the car on our legal and moral judgments, see, e.g., Seo (2019).

¹⁷ There is an ongoing and important literature on intersecting systems of oppression that lies behind my discussion. What are the relationships between racism, sexism, capitalism (etc.)

c. The Social Domain

As mentioned before, the site of ideology critique is *the social domain*. It is difficult to draw a clear line between the social domain and the political domain—and I won't attempt to do so here. But one mark of the political domain is its relationship to the distinctive coercive power of the state, a power that is leveraged in repressive state apparatuses, especially law and its enforcement. The social domain is characteristically structured by norms, expectations, and identities—developed within the ideological state apparatuses and the internalization of the cultural technē—with law serving, in many cases, only as a fallback. So the primary questions for social critique are not the appropriate structure and limits of the state, but rather what practices we should engage in, what social norms we should embrace, and how we should go on, from here, together.

As a result, the normative questions are not primarily whether an agent acts rightly or wrongly, or whether an agent is blameworthy. Nor are the normative questions about what is permissible for the state to regulate and enforce. Rather, the question is whether we (collectively) are warranted in creating, maintaining, or changing a practice or structure. An individual can be treated unjustly *qua* individual by others, or by the state. But within the social domain individuals are vulnerable to perpetrating or suffering injustice by virtue of their social positions. The aim is to improve our social practices and social structures to eliminate this *positional vulnerability*.

A standard strategy for deciding whether a social practice is acceptable is to argue that the practice is in all participants' long-term self-interest, and proposals for change should be evaluated through a kind of collective cost-benefit analysis. In some cases, this answer is straightforward: some practices clearly and systematically deprive individuals of what's necessary for a minimally decent life, or the development of basic capabilities, and these should be changed for the familiar reasons. But there are several reasons why ideology complicates this answer.

in the contemporary social order? See, e.g., Carastathis 2014; Dawson and Katzenstein 2019; Haslanger 2020d; Sewell 2005. The tensions and contradictions between systems are relevant to my response to Celikates' explanatory challenge, as will become clearer below.

First, we cannot judge what is a minimally decent life in the abstract, as Adam Smith's classic comment about linen shirts makes clear.

By necessities I understand not only the commodities which are indispensably necessary for the support of life, but whatever the custom of the country renders it indecent for creditable people, even of the lowest order, to be without. A linen shirt, for example, is, strictly speaking, not a necessary of life...in the present times, through the greater part of Europe, a creditable day-labourer would be ashamed to appear in public without a linen shirt, the want of which would be supposed to denote that disgraceful degree of poverty which, it is presumed, nobody can well fall into without extreme bad conduct. Custom, in the same manner, has rendered leather shoes a necessary of life in England. The poorest creditable person of either sex would be ashamed to appear in public without them. In Scotland, custom has rendered them a necessary of life to the lowest order of men; but not to the same order of women, who may, without any discredit, walk about barefooted...Under necessities, therefore, I comprehend not only those things which nature, but those things which the established rules of decency have rendered necessary to the lowest rank of people.

(Smith 1776/1977, 1168)

Smith's criterion by which to judge what is socially necessary—what is necessary in order to appear in public without shame—is important for its recognition of the social bases of self-respect; but it also reveals the potential for ideological distortion. Why do men need shoes in Scotland, but not women? And similar concerns might be raised about capabilities. Martha Nussbaum includes on her list of basic capabilities, for example, “bodily health” (Nussbaum 2001). But there is much controversy among disability rights activists about the medicalization of disability and the ableism embedded in the dominant contemporary conception of health (Tremain 2001; Barnes 2016). Necessities—not just what we take to be necessities—are ideologically mediated (Fraser 1989a, 1989b).

Second, who, exactly, is the self whose interests are at issue? If human development is shaped to produce socially fluent individuals—the “good citizens” who “work all by themselves”—and if our identities are formed

by reference to a framework of social positions, then are the relevant interests we are aiming to protect the interests of a socially situated self?¹⁸ It would be hard to deny that I—a white cis-woman—have an interest in identifying with and fluently occupying the position of white woman. Failing to do so brings with it substantial costs. But I also have an interest in overturning the local unjust race/gender regime so that race and gender as we know them are no longer imperatives. The problem, more generally, is that the very practices that shape us as social individuals are the ones that function ideologically and so are the targets of critique. But it would also be a mistake to think that the interests in question are those of an unsocialized human being, or a bare self not already embedded in a society and culture.

d. Critique as Emancipatory

Within critical theory, especially Frankfurt School Critical Theory, there is a tradition that insists that a successful ideology critique will be, itself, emancipatory. In his classic work on the Frankfurt School, Raymond Geuss (1981) claims that one of the three “essential distinguishing features of a ‘critical theory’” is that:

1. Critical theories have special standing as guides for human action in that:

(a) they are aimed at producing enlightenment in the agents who hold them, i.e. at enabling those agents to determine what their true interests are;

(b) they are inherently emancipatory, i.e. they free agents from a kind of coercion which is at least partly self-imposed, from self-frustration of conscious human action. (1–2)

Although I agree that a successful ideology critique should aim to provide the resources for individuals in the grip of an ideology to better

¹⁸ I articulate this concern in the language of ‘interests’; if one opts instead for the language of ‘preferences,’ then the problem is even worse.

understand their situation, not all agents will take up these resources or accept the proposed reconceptualization of their social milieu. And of course, thinking alone—or accepting a critical theory—does not free us from coercion, even self-imposed coercion (it may actually increase the need for coercion as we no longer acquiesce to our conditions once we see them for what they are).

As social critics, we should distinguish the *illumination problem*—how we can get others, especially those who are subordinated or oppressed through the working of ideology, to recognize their position and work for social justice—from the *justification problem*. A critique may be justified without providing illumination or emancipation to everyone who encounters it or even understands it.¹⁹ There are multiple determinants of belief other than being given good reasons supporting the claims in question.

We should also distinguish the *justification problem* from the *political problem*. The justification problem concerns whether the critic has a justified complaint against the current social order, i.e., that some practice or set of practices is harmful or unjust. The political problem is what we, collectively, should do about the warranted complaint (and how to decide). As mentioned above, the social critic's role is primarily negative—to identify and analyze forms of injustice that are masked by ideology. It is a further question to determine how the injustices should be remedied. Rarely can all *pro tanto* political complaints be adequately addressed; in general, solutions to collective action problems distribute, but do not eliminate, benefits and burdens. Danielle Allen (2001, 859) argues that “sacrifice makes collective democratic action possible.” She develops this further in her book on citizenship, democracy, and the Civil Rights Movement (2004):

Democracy is not a static end state that achieves the common good by assuring the same benefits or the same level of benefits to everyone, but rather a political practice by which the diverse negative effects of collective political action, and even of just decisions, can be distributed equally, and constantly redistributed over time, on the basis of

¹⁹ This is a point that Shelby (2003, 2014) has made convincingly.

consensual interactions. The hard truth of democracy is that some citizens are always giving things up for others. (29)

In unjust societies, the problem is not that some groups lose out, are outvoted, or suffer substantial costs; this happens in just democracies. Rather, the system is set up so that the pattern of redistribution of costs and benefits over time is skewed: one group makes the sacrifice, another group gets the benefits, over and over and over. For example,

For a long time, in this country, the solution to this paradoxical fact that most democratic citizens are, at the end of the day, relatively powerless sovereigns was the two-pronged citizenship of domination and acquiescence. These old bad habits dealt with the inevitable fact of loss in political life by assigning to one group [White folk] all the work of being sovereign, and to another group [Black folk] most of the work of accepting the significant losses that kept the polity stable. (2004, 41)

I distinguish these various problems—justification, illumination, politics—to set some of them aside. I will not assume that the social critic will be able to convince all others who understand their critique that their critique is warranted, or that critical theory is, in itself, emancipatory. I will not assume that the social critic will have, or must provide, a solution to the problem identified that results in greater justice. My goal is more modest: to sketch one way to produce warranted social critique under conditions of ideology.

4. Methodological/Epistemic Challenge

Celikates' methodological challenge situates us at a skeptical moment: If we, ourselves, may be in the grip of an ideology, how can we judge what is emancipatory? The basic problem of ideology critique is often stated by those in the tradition of the Frankfurt School as what I will call the "critical dilemma." Because Critical Theory aims to motivate and guide social change, it cannot rely on a set of "external" imported values: "any 'strong,' context-transcending form of social criticism necessarily brings

the risk of paternalism or even despotism” (Honneth 2009, 44). Of course, the correlative problem is that if one can only rely on the locally entrenched value horizon, then it is unclear that one will have the resources to break through the grip of ideology (Honneth 2017, 2). One solution—that is designed to disrupt adherence to the existing practices and also avoid paternalism and vanguardism—is to challenge the formal or epistemic workings of the ideology, rather than imposing substantive values from “outside.” So, properly speaking, critique demonstrates that the ideology has epistemic flaws (e.g., is “self-contradictory”) and provides other epistemic resources (and practices) to unmask it, without taking a moral stand (Stahl 2017; Jaeggi 2018). This solution to the critical dilemma is often referred to as “immanent critique.”

Celikates addresses the critical dilemma by casting immanent critique as a “second-order” project that takes the form of “reconstructive critique” (2018, part III). Because ideologies “block the development and/or exercise of the reflexive and critical capacities” of the agents in question,

ideology critique can be understood as second-order critique: If ideologies hide the possibility of criticizing (and transforming) these very ideologies and the problematic first-order phenomena they mask, then the first aim of the critique of ideology has to be to identify these blockades of critique and to work towards their dissolution. In this respect, ideology critique can be seen as taking a procedural turn: Its task is not so much to replace a mistaken or distorted view of social reality with one that is correct (as Althusser implies), or to develop a substantial vision of how society should be organized (as mainstream political philosophy does); rather, its task is to make it possible for agents to ask these questions and collectively look for answers to them themselves. (Celikates 2016, 17)

The critic’s primary goal should be to open space for resistant voices to be heard and allow the community to determine its own collective values and the social practices to further them.

I am sympathetic to Celikates’ proceduralism, and to the fallibility of any such process. I agree that one crucial aim of ideology critique is to

identify and remove the epistemic barriers ideology creates. However, I don't agree that ideology critique should be primarily or exclusively "second-order." There are methods for undertaking ideology critique that yield first-order moral claims. I will sketch some of these methods in the next section, but before doing so, it is worth considering a bit further the set-up for the critical dilemma framed as a choice between internal and external critique.

As I understand the critical dilemma, it rests on assuming a divide between the critical theorist and those directly affected. On the one hand, there are those who are "in the grip" of the ideology and participate in the practice on terms that mask its injustice; on the other hand, there are those who know how ideology works to hide or distort the social conditions but who are not sufficiently embedded in the practice to formulate the moral critique. As Celikates suggests, a crucial commitment of critical theory is to listen to first-person (and first-person plural) knowledge claims of the oppressed. This commitment is partly grounded in epistemic humility: we should listen to those directly affected by the practices in question because they are likely to have better access to morally relevant facts. However, the commitment to listen to those directly affected is sometimes a claim of epistemic *entitlement* by those who are members of such oppositional groups. Why suppose that the critical theorists are not those directly affected? Who are the critical theorists anyway?

One might argue that the task of ideology critique is to address those who are complicit in the unjust practices. The critic may be directly affected but, *qua* critic, is not complicit, and *qua* complicit is not critical. So at the very least, there are two moments in the critical project. The *critic in me* may address my complicit self; and *my complicit self* may respond by rejecting the critique on terms provided by the practice I'm engaged in. In effect, this is to embed the critical dilemma within the psychology of the individual.

I find this move implausible and unhelpful. It presupposes the existence of a critical perspective (the critic "in me") that is part of what we need to explain: How does the critical perspective arise, even within an agent; and how, or under what conditions, is it warranted? Moreover, the supposed tension between the critical and complicit perspectives within

an agent does not do justice to the phenomenology of consciousness raising. Rather, in the process of developing a critical consciousness, the agent undergoes a paradigm shift. This is compatible with being able to “see” the world through both paradigms, but the new paradigm illuminates social reality and brings with it a new sensibility. In doing so, it seems to carry authority. The questions I am asking are: Under what conditions should we trust this shift? When does the new paradigm legitimately carry authority? And can we make warranted claims against others based on the new paradigm, even if they have not embraced it? I’ll return to Celikates’ three challenges below.

5. The Epistemology of Consciousness Raising

a. Case Studies

Under ideological oppression, critique happens in a million ways every day (Scott 1990; Ewick and Silbey 1995, 1999, 2003; Collins 2002; Khader 2011). Some of it is explicit, some not; some of it is warranted, some not; some of it is empowering, some not. And sometimes it builds into a movement. Not all social movements begin in consciousness raising. Especially when repression is regular and obvious, there is often a broad consensus on the injustice and other moral violations; and when there is a longstanding tradition of critique, one can be brought up with a critical consciousness, even as one participates in ideologically shaped practices. However, as discussed above, critical consensus is more difficult to achieve under conditions of ideology.

In what follows, I will consider a particular form of critique that arises in and through a practice that is sometimes called “consciousness raising.”²⁰

²⁰ I am aware that the term ‘consciousness raising’ or ‘CR’ is dated. In my courses, students have teased me that “nowadays we call that ‘raising awareness’.” I want to hold on to the terminology of consciousness raising, however, at least for the time being, because I think there is an epistemic phenomenon worth considering that has been seriously neglected, and the term ‘consciousness’—understood as an oppositional consciousness—has a history that can help us track it. Note that in this tradition (and the related Marxian one), the term ‘consciousness’ is not just a cognitive awareness, but an orientation, a sensibility that brings with it more than just beliefs.

Consciousness raising is a collective activity—done with others—and prompts a paradigm shift in one’s orientation to the world. (See, e.g., Mackinnon 1989, ch. 5; Bartky 1975; Redstockings 1978; Frye 1990; Crow 2000; McWeeny 2016; Crary 2015; Toole 2019.) This includes a shift in what facts become accessible, our interpretation of them, and what responses are called for. It is not easily reversed. The experience of such a paradigm shift is powerful, but its adequacy or warrant is not guaranteed. If a movement is to be built on such a paradigm shift, and if movements are to make warranted claims against others, then we need to think more about the conditions under which consciousness raising provides knowledge, and what sort of knowledge it provides. In the next sections, I will provide a sketch of some of the main features of an epistemology of consciousness raising, as I see it.²¹ There is much that needs further discussion and elaboration. I start with a brief description of two examples.

Combahee River Collective (1983) (“A Black Feminist Statement”)

In 1974 a group of Black women started meeting in response to their experiences in everyday life and in the Civil Rights Movement (CRM) and the Women’s Liberation Movement (WLM).²² Their frustration had roots in their situation: “the political realization that comes from the seemingly personal experiences of individual Black women’s lives” (266), and also the failures of both the CRM and the WLM to give them the tools to develop an adequate response: “there was no way of conceptualizing what was so apparent to us, what we *knew* was really happening” (266). Through a process of consciousness raising, they explored the cultural and political dimensions of their experience, and developed new terms and concepts. For example,

²¹ It has been difficult to find literature in philosophy on the epistemology of consciousness raising, which, I suppose, is not surprising. I am anxious in writing this section because the phenomenon is huge and multi-faceted, and has a meaningful history, with both strengths and pitfalls. I am vividly aware that my research has not been thorough, but in an effort to draw attention to the phenomenon and encourage others to work on it, I am offering what I have managed to put together thus far.

²² To find the published statement along with important interviews with some of the authors, see Taylor (2017).

We discovered that all of us, because we were “smart,” had also been considered “ugly,” i.e., “smart-ugly.” “Smart-ugly” crystalized the way in which most of us had been forced to develop our intellects at great cost to our “social” lives. (268)

Through CR, they reached the “shared belief that Black Women are inherently valuable, that our liberation is a necessity not as an adjunct to somebody else’s but because of our need as human persons for autonomy” (33) and “to be recognized as human, levelly human, is enough” (267).

The group that persisted through 1977—when the statement was written—decided that CR was not enough. They developed a study group, and decided to promote their cause through writing, publishing, lecturing, and other activist organizing. They conclude,

We believe in collective process and a non-hierarchical distribution of power within our own group and in our vision of a revolutionary society. We are committed to a continual examination of our politics as they develop through criticism and self-criticism as an essential aspect of our practice. (273)

“The Girls Fought Back.”

On March 26, 2019, the *Washington Post* published an article about a group of girls at Bethesda-Chevy Chase High School who learned that their male peers had created a “list” that “ranked and rated [them] based on their looks from 5.5–9.4, with decimal points to the hundredth place.”²³ This kind of activity is not new at the particular high school and occurs virtually everywhere in some form or another. But a subset of the girls on the list, inspired by what they had learned through the #MeToo movement, were upset and complained to the principal. One male student was given detention and that was supposed to be the end of it.

But one of the girls, Nicky Schmidt, texted with her friends after the disciplinary action was announced and they called on others in the International Baccalaureate (IB) program to meet at the main office the

²³ <https://www.washingtonpost.com/lifestyle/2019/03/26/teen-boys-rated-their-female-classmates-based-looks-girls-fought-back>.

next day to protest the inadequacy of the school's response. Forty girls showed up. As a result, the school hosted a 2.5-hour discussion with all students, including those who produced the list. At this meeting, "Several girls delivered personal and impassioned speeches describing not only their presence on the list but also their previous experiences with sexual abuse, harassment and objectification, both inside the school and outside of it." After this meeting, the boy responsible for the list said, "When you have a culture where it's just normal to talk about that, I guess making a list about it doesn't seem like such a terrible thing to do... It's easy for me to lose sight of the consequences of my actions and kind of feel like I'm above something... [But] It's just a different time and things really do need to change." Collective action was then planned to implement policies and practices aimed to reduce similar behavior in the future.

b. Sources of Oppositional Consciousness

Jane Mansbridge uses the term 'oppositional consciousness' to capture a particular kind of response to oppression. She suggests (drawing on Foucault) that oppositional consciousness *in liberation movements* (cf. social responsibility movements such as environmentalism or the peace movement) requires:

a gut refusal to be subordinated rooted somewhere in every human being... To form an effective basis for collective action, gut refusals need cognitive and emotional organizing. They need an injustice frame ... They need an apparatus involving both reason and emotion.

(2001, 4)

Iris Young suggests that resistance begins with a "desiring negation" (1990, 6–7):

Desire... creates the distance, the negation, that opens the space for criticism of what is. This critical distance does not occur on the basis of some previously discovered rational ideas of the good and the just. On

the contrary, the ideas of the good and the just arise from the desiring negation that action brings to what is given.

Each social reality presents its own unrealized possibilities, experienced as lacks and desires. Norms and ideals arise from the yearning that is an expression of freedom: it does not have to be this way, it could be otherwise.

Many feminist and anti-racist theorists have argued that a critical perspective emerges from participation in multiple social “worlds” that are not in sync. For example, Du Bois coined the phrase “double consciousness” (1997/1903); Patricia Hill Collins describes the position of the “outsider within” (1989); Gloria Anzaldúa points to the experience of the *mestizaje* (1987); Maria Lugones describes “world traveling” (1987; also Fanon (1967/1952); Ellison (1972/1952); hooks (1995); Scheman (1997); and many others). These authors provide vivid and profound examples, but it is important to note that we are all positioned in multiple social worlds and are asked to navigate between them, e.g., the worlds of work and of home, the worlds of family of origin and of family of choice, the worlds of adults and of children (as a parent or teacher), the worlds of the market and of personal, family, or religious life. Transitions between these worlds can happen seamlessly, but often prompt a sense of dislocation, demand adjustment, and provide opportunities for critical reflection (or at least a “desiring negation”).

Drawing on empirical case studies, Mansbridge and Morris (2001, 5) argue that certain tools are valuable in moving from an impulse to resist to an “injustice frame.”

An existing *oppositional culture* provides ideas, rituals, and long-standing patterns of interaction that overt political struggle can refine and develop to create a more mature oppositional consciousness . . . a history of *segregation* with some autonomy, providing “free spaces” for the elaboration and testing of ideas; *borrowing* from previous successful movements; the *synthesis* of more than one oppositional strand, creating more than the sum of its parts; mutually supportive

interaction, bridging divides in emotional commitments; and *consensus creativity* by activists, drawing on the traditions and practices of everyday life. (Mansbridge and Morris 2001, 7–8)

Note that the tools Mansbridge provides are social resources and practices that enable one to move beyond one's own individual frustrations to a collective and historically rooted understanding of the situation. Oppositional consciousness transforms into a movement when those in the group "demand changes in the polity, economy or society to rectify those injustices" (Mansbridge and Morris 2001, 1).

c. Epistemic Credentials

As argued above, a primary task of social critique under conditions of ideology is to articulate a warranted moral claim *in the name of a subordinate group or groups*. The claim is made against those with whom one coordinates—in a classroom, a family, an institution (workplace, civic organization), a nation—and makes a demand that the terms of coordination be changed. I assume that one need not be a member of the subordinate group in order to demand justice with them (Pohlhaus 2002). For example, as a White person, I can have a "gut refusal" to participate in racist practices, and my own experience of dislocation between racist and anti-racist contexts enables me to envision more just possibilities and create opportunities for change. In the example of Bethesda-Chevy Chase High School, the process enabled (some of) the boys who made the list to gain consciousness of a sexist culture. But in a liberation movement, the process of articulating a claim through consciousness raising typically begins with those directly affected.²⁴

Oppositional consciousness arises and can be justified in a variety of ways; I am focusing on what I take to be hard cases where there isn't a critical paradigm in place, where ideological oppression has interpellated

²⁴ Much more should be said here about the different forms of consciousness raising and the different kinds of participants. In the description that follows, I'm assuming that we are starting with those "in the grip" of ideology, not guided by a "theorist" who already has escaped the grip.

individuals to be “good subjects” in a variety of practices. The process I am exploring involves a resistant reaction that can evolve into a complaint, and may result in a *pro tanto* moral claim.²⁵ I use bullets rather than numbers below because the sequence of steps may not always occur in the order presented.

- There is a moral “gut refusal” to comply with or accept a practice, a “desiring negation” that yearns for and imagines other possibilities. Such a refusal may simply be a personal indication of displeasure, a whine, but does not rise to the level of a complaint against others. How do we transform whining or displeasure into a proper complaint?

I am assuming, for the purposes of the discussion, that whining just expresses a negative preference, a preference against something. Whining, in the sense I mean, is not characterized by tone of voice, but is characterized by failing to even provide a *prima facie* reason for others to act differently.²⁶ A child may whine when asked to go to bed, but this is not, even on the face of it, a reason for a parent to adjust bedtime.

Moreover, individual displeasure is not usually, by itself, sufficient evidence that there is a positional vulnerability. The wrong or harm may be personal: it may be that the individual has been selected for mistreatment, but not on the basis of group membership. From the point of view of members of subordinate groups, this can be a moment of ambiguity:²⁷ Is it something about me (as an individual), or is this person really a sexist/racist/...? Am I in the wrong, or are they in the wrong? For those in the grip of ideology, it can be tempting to resolve the ambiguity by concluding that they have failed to live up to a set of presumed standards. A crucial moment in consciousness raising occurs when one gains evidence that the negative behavior is systematic,

²⁵ In this section, I draw significantly on Anderson’s pragmatism (2014, forthcoming).

²⁶ I mean *prima facie* and not *pro tanto*. The issue is whether, on the face of it, there is reason to think that the preference is one that should override the default. Some *prima facie* reasons do not, after further consideration, rise to the level of *pro tanto* reasons.

²⁷ Walton and Brady (2017) discuss several related phenomena, including attributional ambiguity, social-identity threat, and belonging uncertainty, all of which are subject to intensification because of a kind of looping effect (which they call ‘recursion’).

group-based, and unwarranted and, most importantly, realizes that they are not the problem.

Another explanation of the mistreatment may be that one is dealing with a bad actor rather than a systematic phenomenon. A bad actor may have problematic, but idiosyncratic, responses to individuals in a particular group; but this is not the phenomenon social critique seeks to address. For example, if an individual fails to take seriously people who wear baseball caps—due, perhaps, to a personal association formed long ago—it is important to become aware of this. This individual's bias may become apparent in the context of consciousness raising, but the goal of consciousness raising is to identify systematic and structural vulnerability and provide a critique of social practices rather than individuals. There are, of course, cases in which an individual bias is caused by broader structural phenomena—perhaps the clothing in question is a marker of social (racial, ethnic, class) status—and the individual's behavior is a symptom of a systematic problem. These are the sorts of examples that social critique can build on.

Because the inquiry is into social—structurally produced—injustices, and because it is difficult as an individual to determine what the social patterns are and how to interpret them (note that being a target of negative behavior can give rise to shame and reluctance to share the experience with others), the process from here forward is collective. This is not to say that an individual cannot, working alone, identify systematic injustice and articulate a warranted critique. But the method employed by such an individual would not be, strictly speaking, consciousness raising; and in order to move forward as part of a movement, the critique would have to give rise somehow to a broader, shared, paradigm shift. So an essential part of consciousness raising is group participation.

- Articulate the concern to others within the same (affected) social group; test the reaction against the experience of others. Consider: Is the problem individual or social? Am I over-reacting? Are others treating me this way because I am acting badly? Is the agent simply a bad actor? Is this occurring because of a *positional vulnerability*?
 - o To achieve this, it is often important to create counter-publics where the subordinated can complain to each other without

being “corrected” by members of the dominant group, where they can be heard. (Mansbridge and Morris 2001, 7–8; Fraser 1990; Dotson 2011; Dotson 2014.)

- o The “testing” process—both articulating the concern and responding to it—should involve forms of *bias reduction* and *consideration of epistemic injustice* of all sorts. Testimonial injustice and gaslighting are serious risks. As Elizabeth Anderson (forthcoming, 7) notes, there is compelling empirical evidence of systematic power biases: “Standing in a position of superior power over others tends to bias the moral sentiments of the powerful, in at least three ways: it reduces their compassion, activates their arrogance, and leads them to objectify subordinates.” Because of the diversity and power differentials within subordinated groups, one may need to narrow one’s community in order to adequately resolve whether there is a positional vulnerability and to identify the particular social position that renders one vulnerable to the harm or wrong in question.
- o The process allows for, even encourages, hermeneutical invention. Individuals within the group can sometimes rely on existing identities, but in other cases new “identities” are called for (Mansbridge and Morris 2001, 9). Shared identities (Black feminist, queer) allow for a cultivation of trust, new language, shared interests, etc. Patterns can then become more visible and new hermeneutic resources developed (‘smart-ugly,’ ‘White fragility,’ ‘mansplaining,’ ‘himpathy’ (Manne 2017)).

A distinction feature of consciousness raising is that it involves trying on different perspectives, vocabularies, sensibilities, to notice facts that have been occluded—empirical facts, morally relevant facts, facts about possibilities. Shifts in orientation can be prompted by historical inquiry, the idiosyncratic and creative suggestions by individuals, existing oppositional cultures (#MeToo), local narrative traditions, or comparisons made possible by participation in different practical domains, e.g., work/home. One of the most effective tools of ideology is the systematic maintenance of ignorance (Mills 2007). Serene Khader makes this point in relation to Western normative hubris:

the idea that some people are uniquely situated to bring about moral progress draws on a certain way of construing the non-normative facts about the world. Indeed, one of the major contributions of decolonial, postcolonial, and transnational feminist theories has been to excavate the nonnormative (or not directly normative) ways of seeing that make imperialism seem legitimate, and even necessary. Missionary feminists take the West to be an agent of morality, and they preserve the deep psychological and ideological association between the West and morality by filtering away information that might reflect poorly on the West or its values. (2019, 31)

- o Develop and experiment with a new paradigm. This includes a new way of thinking, but is broader than a cognitive shift (Railton 2014). It includes new practices that call for different ways of interacting, different relationships, different affective responses. Counter-publics are another site for such experiments in living (Anderson 1991).
- *Develop a hypothesis* about the forms and causes of structural injustice—the particular practices that entrench subordination and block change. (See also Mansbridge and Morris 2001, 5.) This may involve targeting injustices in various domains: epistemic, material, legal, cultural.

For example, the feminist movement has challenged legal responses to marital rape and domestic violence, the material demands of women's "second shift," the wage gap, rape culture, and multiple forms of testimonial injustice. I hope there is no need to list the impressive and ongoing movement work calling attention to unjust practices that create positional vulnerability on the basis of class, race, sex/gender, disability, LGBTQ+ status, immigration, etc.

However, not all paradigm shifts lead to greater justice, and not all hypotheses about the causes of injustice are warranted. For example, Anti-Vaxxers believe that vaccines harm their children; Neo-Nazis believe that Jews (and other minority groups, including LGBTQ+) are

inferior “by nature” and responsible for degrading European culture. Consciousness-raising groups are, by design, focused on the experiences and interpretations of those who participate in the group. But the evidence available to them is limited and efforts must be made to draw on whatever empirical knowledge is obtainable.

- *Test the hypothesis.* Is the hypothesis generated from within the new paradigm empirically adequate? Is the hypothesis the best explanation of the injustices? Draw on critical social science. Revise the hypothesis, as needed.

Eric Olin Wright describes an emancipatory, or critical, social science:

It is not enough to show that people suffer in the world in which we live or that there are enormous inequalities in the extent to which people live flourishing lives. A scientific emancipatory theory must show that the explanation for this suffering and inequality lies in specific properties of institutions and social structures. The first task of emancipatory social science, therefore, is the diagnosis and critique of the causal processes that generate these harms. (Wright 2010, 11)

This is a moment when the distinction between the project of illumination and the project of justification plays a crucial role. It is unlikely that Neo-Nazis or Anti-Vaxxers are going to be convinced by the empirical evidence that their hypotheses are false. This does not mean, however, that it is inappropriate for us to regard their challenges as misguided. Of course, in order to maintain political order and protect certain speech rights, we may be required to tolerate the expression of their views. But a demand that a practice be changed based on empirically refutable claims, or claims with false presuppositions, is not one that needs to be honored.

One problem, of course, is that the project of empirical justification and refutation is more complicated than I have thus far suggested. Some of the claims made by misguided oppositional groups may not be empirically refutable. Existing social science may not be prepared to adopt the hermeneutical resources that arise through the process of consciousness raising (though this is part of the remit of critical social

science); and some of the claims may be irreducibly normative and constitute a fundamental moral disagreement (though see Moody-Adams 1997). But the process of epistemic validation is not foundationalist. The best that any inquiry—empirical or not—can achieve is a holistic balancing of considerations. And scientific inquiry has managed to weather paradigm shifts before without giving up all standards (Kuhn 1962).

Part of what's at issue is the interpretation of reality: what matters and what doesn't, what is parallel to what, what narrative threads tie things together (Scheman 2017; Walker 2007; Crary 2009). Our modes of evaluation rely on a broad range of human capacities, not just the concepts that contemporary moral theory offers. Ideology can truncate these capacities; it can also teach us which to trust and which not. The aim of consciousness raising is not to reach certainty or to offer evidence that would be compelling to all who consider it. The task is to engage in epistemically responsible practices that push us beyond what is taken to be common sense, while also affording some degree of objectivity.²⁸ After testing and revisions of our hypothesis have reached a stable point, we—i.e., those undertaking consciousness raising—move to the moral claim.

- Articulate a *moral claim* challenging the practice, e.g., this (part of the) practice is unjust, oppressive, harmful, or wrongful.

On this view, an oppositional consciousness is warranted insofar as it moves from a “gut refusal” to a moral claim through a collective examination of shared experience that is guided by sound epistemic norms. What norms are “sound” is not simply a matter of what the dominant culture recommends, but should be guided by best practices of social

²⁸ Work on objectivity within critical theory (feminist, anti-racist, post-colonial) is extensive and challenges many of the traditional assumptions about what objectivity consists in. However, it does not, for the most part, reject the value of objectivity tout court. For example, the feminist empiricist literature, e.g., Longino 1990, recommends a procedural account whereby only communities that meet certain standards of diversity and critical engagement can count as objective. Generally, the goal is to provide a conception of objectivity that allows for values to play a legitimate role. Although my argument is seriously incomplete unless and until I provide some guidance on the criteria for objectivity, this gesture is the most I can accomplish in this chapter.

psychology, empirical investigation, critical epistemologies, and the lived experience of those in the subordinate group. The resulting claim is made on behalf of a social group and warranted through their collective efforts. Although, as Celikates argues, changes to the epistemic practices are required in order to loosen the grip of ideology, critique sometimes emerges in the collective response to one's situation through consciousness raising, and yields first-order normative claims.

An oppositional moral claim is not, simply by virtue of being the result of such a process, dispositive. It offers a *pro tanto* consideration for collective deliberation in a process of contentious politics. I take no stand here on the conditions for the legitimacy of the political process from this point forward.

6. Conclusion: The Normative Basis for Ideology Critique

Under conditions of ideology there is, by hypothesis, a range of social practices that oppress a group; however, some do not experience them as oppressive. Celikates raised three challenges for the possibility of ideology critique. We are now in a position to consider responses (I repeat the challenges from above).

(i) *Normative challenge*: What makes an ideology problematic? Are there objective moral truths by reference to which we can judge a social arrangement defective or unjust? If so, how do we gain knowledge of those truths?

There are moral truths about the injustice of particular historically specific practices and structures. The Atlantic slave trade was wrong. Nazi genocide was wrong. The current systematic violation of African-American civil rights is wrong. The global sex trade is wrong. These are not truths we learn from theory; theorizing is guided by these truths. And we know these things. These are not the controversial cases in which moral knowledge is at issue. More importantly for our purposes, these are not cases of ideological oppression; they are examples of grotesque repression. The challenge for ideology critique is not to demonstrate the obvious.

The difficult questions concern the critique of what we learn through being interpellated into practices, what becomes common sense to “good citizens.” Think of the organization of capitalist society and assumptions concerning wage contracts, “right to work” slogans, the division of labor in the family, and such. In the process of consciousness raising, we develop an alternative description and explanation of a phenomenon that reveals morally relevant aspects that ideology masks. Once these aspects are revealed, or diagnosed, the phenomenon is no longer viewed as innocent or as commonly represented. For example, once one sees wage labor as the extraction of value from workers that is pocketed as profit by capitalists, i.e., once one sees capitalism as founded on exploitation, one cannot regard capitalism as a benign economic system. Our gaze shifts so that we find parallels between cases that horrify us and ones we take for granted, e.g., “wage slave,” “private government” (Anderson 2017). Whether or not the parallels stand up to scrutiny is an open question. But if the parallels are sufficiently strong, or if we agree that the new interpretation better guides our practice, then we are entitled—epistemically and morally—to make a claim on its basis.

(ii) *Methodological or epistemological challenge*: From what standpoint does the critic speak?: “which insights of which agents—given that they usually do not constitute a homogeneous category—the critical theorist articulates” (2016, 4).

The critical theorist that I have described is embedded in a movement. She is not an “outsider” who is trying to convince the subordinate to rise up by providing them a theory. She is engaged with others in consciousness raising, and is articulating the insights that come from participation in it together. The claims that arise from the movement may not be ones that all members of the subordinated group support. But this does not show that the critique is misguided. It may be that the values the resistant rely on when making claims of being harmed or wronged are at odds with what others engaged in the practice value. But that does not delegitimize their claims. Social practices are cooperative enterprises, and if parties to the cooperation have reason to think that they are being treated unjustly, or their values are being undermined, there is a *pro tanto* reason for all parties involved to reconsider the practice.

(iii) *Explanatory challenge*: If an ideology functions at the level of a system, “what exactly holds the rather broad conglomeration of partly psychological, partly social mechanisms—from implicit biases via stereotypes to looping effects—together and makes them into elements of one ideology”? And how does critique disrupt the systematic injustice sustained by the ideology?

I don’t agree with the assumption that “ideology functions at the level of a system.” Although it is a common theme in Frankfurt School Critical Theory that the target of ideology critique is the socio-historical “totality,” this is not actually true of many critiques that arise in the context of social movements. I do agree that social justice requires systematic change, but I see no reason to think that unjust systems are sustained by a single ideology that is the proper object of critique; on my view, ideologies are a collection of social meanings that may well materially or accidentally coincide to produce the problematic effects. This fragmentation is not a bug, but a feature of my view, since it also allows us to find fissures for leveraging critique.

Celikates asks a further question, however: How do the multiple mechanisms of systematic injustice work together to sustain it? This is an important question for deciding how to intervene in systems in order to change them. However, I don’t believe that an account of ideology critique requires an answer to this question, or that requires us to provide identity conditions ideologies so that we can differentiate them.

I began the chapter by asking what entitles us to claim that a cultural technē—the set of social meanings that shape our practices—is ideological; in other words, how, under conditions of ideology, can we establish a basis for social critique? I’ve argued that given both the epistemic challenges posed by ideology and the historically situated task of challenging social practices, we should look for methods other than developing an ideal theory of justice to undertake critique. I have suggested that there are multiple ways of pursuing critique, and have described the process of consciousness raising as a method that yields an alternative paradigm for understanding and engaging in social life. Although I have not given a full defense of CR as a basis for normative knowledge, I’ve pointed to some of its epistemic credentials that can warrant groups to make *pro tanto* claims on others concerning the

injustice of shared practices. Very broadly, I hope I have opened up some space within moral inquiry to consider the pernicious effects of ideology, not only on the social systems we embody but also on our theorizing, and also provided some resources from critical theory to think together about how we might go on.²⁹

Works Cited

- Agarwal, Bina. 1992. "The Gender and Environment Debate: Lessons from India." *Feminist Studies* 18(1): 119–158.
- Agarwal, Bina. 2009. "Rule-Making in Community Forestry Institutions: The Difference Women Make." *Ecological Economics* 68: 2296–2308.
- Allen, Danielle. 2001. "Law's Necessary Forcefulness: Ralph Ellison vs. Hannah Arendt on the Battle of Little Rock." *Oklahoma City University Law Review* 26: 857–895.
- Allen, Danielle. 2004. *Talking to Strangers: Anxieties of Citizenship Since Brown v. Board of Education*. Chicago, IL: University of Chicago Press.
- Althusser, Louis. 2014/1971. "Ideology and Ideological State Apparatuses." In *On the Reproduction of Capitalism: Ideology and Ideological State Apparatuses*. Trans. Ben Brewster. London: Verso.
- Anderson, Elizabeth. 1991. "John Stuart Mill and Experiments in Living." *Ethics* 102(1): 4–26.
- Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.
- Anderson, Elizabeth. 2014. "Social Movements, Experiments in Living and Moral Progress: Case Studies from Britain's Abolition of Slavery." *The*

²⁹ Thanks for helpful discussion to: Louise Antony, Ásta, Emad Atiq, Nancy Bauer, Marion Boulicault, Susan Brison, Robin Celikates, Stephanie Collins, Alice Cray, David Estlund, Sahar Heydari Fard, Ann Ferguson, Lorna Finlayson, Rainer Forst, Robert Gooding-Williams, Stefan Gosepath, Elizabeth Hackett, Rahel Jaeggi, Daniel James, Tamara Jugov, Timo Jütten, Serene Khader, Alison Koslow, Quill Kukla, Kristina Lepold, Andrei Marmour, Mirjam Müller, Gaile Pohlhaus, Jr., Jonathan Quong, Tully Rector, Gideon Rosen, Naomi Scheman, Tommie Shelby, Titus Stahl, Katya Stoppenbrink, Jeffrey Stout, Eva von Redecker, Stephen Yablo, Robin Zheng, and many others, including the participants at the Oxford Studies in Political Philosophy Conference (August 2019), and anonymous reviewers for *Oxford Studies in Political Philosophy*. Thanks especially also to David Sobel for his insight, patience, and support.

- Lindley Lecture, University of Kansas. https://kuscholarworks.ku.edu/bitstream/handle/1808/14787/Anderson_Social_Movements.pdf.
- Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (And Why We Don't Talk About It)*. Princeton, NJ: Princeton University Press.
- Anderson, Elizabeth. Forthcoming. "How to Be a Pragmatist." In *Oxford Handbook of Practical Reason*. Edited by Ruth Chang and Kurt Sylvan. Oxford: Oxford University Press.
- Anzaldúa, Gloria. 1987. *Borderlands/La Frontera*. San Francisco, CA: Aunt Lute Books.
- Balkin, J.M. 1998. *Cultural Software: A Theory of Ideology*. New Haven, CT: Yale University Press.
- Barnes, Elizabeth. 2016. *The Minority Body*. Oxford: Oxford University Press.
- Bartky, Sandra Lee. 1975. "Towards a Phenomenology of Feminist Consciousness." *Social Theory and Practice* 3(4): 425–439.
- Bartky, Sandra Lee. 1990. "Foucault, Femininity, and the Modernization of Patriarchal Power." In *Femininity and Domination*. New York: Routledge, pp. 63–82.
- Brown, Les. 1974. "F.C.C. Chief Bids TV Adopt Joint Adult-Show Policy." *The New York Times*, November 26, p. 78.
- Carastathis, Anna. 2014. "The Concept of Intersectionality in Feminist Theory." *Philosophy Compass* 9(5): 304–314.
- Celikates, Robin. 2016. "Beyond the Critical Theorists' Nightmare: Epistemic Injustice, Looping Effects, and Ideology Critique." Presentation at the Workshop for Gender and Philosophy, MIT. May 12.
- Celikates, Robin. 2018. *Critique as Social Practice: Critical Theory and Social Self-Understanding*. London: Rowman and Littlefield International.
- Collins, Patricia Hill. 1989. "The Social Construction of Black Feminist Thought." *Signs* 14(4): 745–773.
- Collins, Patricia Hill. 1991. "Learning from the Outsider Within." In *Beyond Methodology: Feminist Scholarship as Lived Research*. Edited by Mary Margaret Fonow and Judith A. Cook. Bloomington, IN: Indiana University Press.
- Collins, Patricia Hill. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

- Combahee River Collective. 1983. "A Black Feminist Statement." In *Home Girls: A Black Feminist Anthology*. Edited by Barbara Smith. New York: Kitchen Table Press, pp. 264–274.
- Crary, Alice. 2009. *Beyond Moral Judgment*. Cambridge, MA: Harvard University Press.
- Crary, Alice. 2015. "Feminist Thought and Rational Authority: Getting Things in Perspective." *New Literary History* 46(2): 287–308.
- Crow, Barbara A. 2000. *Radical Feminism: A Documentary Reader*. New York: NYU Press.
- Dawson, Michael C. and Emily A. Katzenstein. 2019. "Articulated Darkness: White Supremacy, Patriarchy, and Capitalism in Shelby's *Dark Ghettos*." *Journal of Political Philosophy* 27(2): 252–268.
- Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26(2): 236–257.
- Dotson, Kristie. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28(2): 115–138.
- Du Bois, W.E.B. 1997/1903. "Of Our Spiritual Strivings." In *Souls of Black Folk*. Edited by David W. Blight and Robert Gooding-Williams. New York: Bedford/St. Martin's, ch. 1, pp. 37–44.
- Dworkin, Ronald. 1977. *Taking Rights Seriously*. Cambridge, MA: Harvard University Press.
- Ellison, Ralph. 1972/1952. *Invisible Man*. New York: Vintage.
- Ewick, Patricia and Susan S. Silbey. 1995. "Subversive Stories and Hegemonic Tales: Toward a Sociology of Narrative." *Law and Society Review* 29(2): 197–226.
- Ewick, Patricia and Susan S. Silbey. 1999. "Common Knowledge and Ideological Critique: The Significance of Knowing That the 'Haves' Come Out Ahead." *Law & Society Review* 33(4): 1025–1041.
- Ewick, Patricia and Susan S. Silbey. 2003. "Narrating Social Structure: Stories of Resistance to Legal Authority." *American Journal of Sociology* 108(6): 1328–1372.
- Fanon, Frantz. 1967/1952. *Black Skin/White Masks*. Trans. Charles Lam Markmann. London: Pluto Press.
- Foucault, Michel. 1979. *Discipline and Punish*. Trans. Alan Sheridan. New York: Vintage Books.
- Fraser, Nancy. 1989a. "Women, Welfare, and the Politics of Need Interpretation." In *Unruly Practices: Power, Discourse, and Gender in*

- Contemporary Social Theory*. Minneapolis, MN: University of Minnesota Press, pp. 144–160.
- Fraser, Nancy. 1989b. “Struggle Over Needs: Outline of a Socialist-Feminist Critical Theory of Late Capitalist Political Culture. In *Unruly Practices: Power, Discourse, and Gender in Contemporary Social Theory*. Minneapolis, MN: University of Minnesota Press, pp. 160–190.
- Fraser, Nancy. 1990. “Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy.” *Social Text* 25/26: 56–80.
- Frye, Marilyn. 1990. “The Possibility of Feminist Theory.” In *Theoretical Perspectives on Sexual Difference*. Edited by Deborah L. Rhode. New Haven, CT: Yale University Press, pp. 174–184.
- Geuss, Raymond. 1981. *The Idea of a Critical Theory: Habermas and the Frankfurt School*. Cambridge: Cambridge University Press.
- Hacking, Ian. 2002. *Historical Ontology*. Cambridge, MA: Harvard University Press.
- Haslanger, Sally. Forthcoming. “Methods of Social Critique” In *Publications of the Austrian Ludwig Wittgenstein Society* 28. Edited by Österreichische Ludwig-Wittgenstein-Gesellschaft.
- Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. New York: Oxford University Press.
- Haslanger, Sally. 2017a. *Critical Theory and Practice*. Assen: Koninklijke Van Gorcum.
- Haslanger, Sally. 2017b. “Racism, Ideology, and Social Movements.” *Res Philosophica* 94(1): 1–22.
- Haslanger, Sally. 2018. “What Is a Social Practice?” *Royal Institute of Philosophy Supplements* 82: 231–247.
- Haslanger, Sally. 2019a. “Disciplined Bodies and Ideology Critique.” *Glass Bead*. Site 2: Dark Room, Somatic Reason and Synthetic Eros. <https://www.glass-bead.org/article/disciplined-bodies-and-ideology-critique/?lang=enview>.
- Haslanger, Sally. 2019b. “Tracing the Socio-Political Reality of Race.” In *What Is Race? Four Philosophical Views*. Edited by Joshua Glasgow, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer. New York: Oxford University Press.
- Haslanger, Sally. 2020a. “How Not to Change the Subject.” In *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*. Edited by Teresa Marques and Åsa Wikforss. Oxford: Oxford University Press, pp. 235–259.

- Haslanger, Sally. 2020b. "Taking a Stand: Second-Order Social Pathologies or First-Order Critique." In *Debating Critical Theory: Engagements with Axel Honneth*. Edited by Julia Christ, Kristina Lepold, Daniel Loick, and Titus Stahl. Lanham, MD: Rowman and Littlefield, pp. 35–50.
- Haslanger, Sally. 2020c. "Going On, Not in the Same Way." In *Conceptual Engineering and Conceptual Ethics*. Edited by Alexis Burgess, Herman Cappelen, and David Plunkett. New York: Oxford University Press, pp. 230–260.
- Haslanger, Sally. 2020d. "Why I Don't Believe in Patriarchy." *Philosophy and Phenomenological Research* 100(4): 220–229.
- Honneth, Axel. 2009. *Pathologies of Reason: On the Legacy of Critical Theory*. Trans. James Ingram and others. New York: Columbia University Press.
- Honneth, Axel. 2017. "Is There an Emancipatory Interest? An Attempt to Answer Critical Theory's Most Fundamental Question." *European Journal of Philosophy* 25: 908–920.
- hooks, bell. 1995. *Killing Rage: Ending Racism*. New York: Henry Holt & Co.
- Intemann, Kristen. 2010. "25 Years of Feminist Empiricism and Standpoint Theory: Where Are We Now?" *Hypatia* 25(4): 778–796.
- Jaeggi, Rahel. 2018. *Critique of Forms of Life*. Trans. Ciaran Cronin. Cambridge, MA: Harvard University Press.
- Khader, Serene J. 2011. *Adaptive Preferences and Women's Empowerment*. Oxford: Oxford University Press.
- Khader, Serene J. 2019. *Decolonizing Universalism: A Transnational Feminist Ethic*. Oxford: Oxford University Press.
- Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Kymlicka, Will. 2002. *Contemporary Political Philosophy: An Introduction*. 2nd edition. Oxford: Oxford University Press.
- Longino, Helen. 1990. *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.
- Lugones, Maria. 1987. "Playfulness, 'World Traveling' and Loving Perception." *Hypatia* 2(2): 3–19.
- MacKinnon, Catharine. 1989. *Towards a Feminist Theory of the State*. Cambridge, MA: Harvard University Press.
- Manne, Kate. 2017. *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press.

- Mansbridge, Jane and Aldon Morris (eds.). 2001. *Oppositional Consciousness: The Subjective Roots of Social Protest*. Chicago, IL: University of Chicago Press.
- McGeer, Victoria. 2007. "The Regulative Dimension of Folk Psychology." In *Folk Psychology Re-Assessed*. Edited by Daniel D. Hutto and Matthew Ratcliffe. New York: Springer.
- McWeeny, Jennifer. 2016. "Varieties of Consciousness under Oppression." In *Phenomenology and the Political*. Edited by S. West Gurley and Geoff Pfeifer. London: Rowman and Littlefield International, pp. 149–163.
- Mills, Charles. 2005. "Ideal Theory as Ideology." *Hypatia* 20 (3): 165–184.
- Mills, Charles. 2007. "White Ignorance." In *Race and Epistemologies of Ignorance*. Edited by Shannon Sullivan and Nancy Tuana. Albany, NY: SUNY Press.
- Moi, Toril. 2015. "Thinking Through Examples: What Ordinary Language Philosophy Can Do for Feminist Theory." *New Literary History* 46(2): 191–216.
- Moody-Adams, Michele M. 1997. *Fieldwork in Familiar Places: Morality, Culture, and Philosophy*. Cambridge, MA: Harvard University Press.
- Nussbaum, Martha C. 2001. *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Pohlhaus, Gaile, Jr. 2002. "Knowing Communities: An Investigation of Harding's Standpoint Epistemology." *Social Epistemology* 16(3): 283–293.
- Pollock, John and Joseph Cruz. 1999. *Contemporary Theories of Knowledge*. 2nd edition. Lanham, MD: Rowman and Littlefield.
- Railton, Peter. 2014. "The Affective Dog and its Rational Tale: Intuition and Attunement." *Ethics* 124(4): 813–859.
- Redstockings of the Women's Liberation Movement. 1978. *Feminist Revolution: An Abridged Edition with Additional Writings*. Edited by Redstockings of the Women's Liberation Movement. New York: Random House.
- Sarachild, Kathie. 1978. "Consciousness Raising: A Radical Weapon." In *Feminist Revolution: An Abridged Edition with Additional Writings*. Edited by Redstockings of the Women's Liberation Movement. New York: Random House, pp. 144–150. <https://www.rapereliefshelter.bc.ca/>

- sites/default/files/imce/Feminist-Revolution-Consciousness-Raising—A-Radical-Weapon-Kathie-Sarachild.pdf.
- Scanlon, T. M. 2003. "Rawls on Justification." In *The Cambridge Companion to Rawls*. Edited by Samuel Freeman. Cambridge: Cambridge University Press, pp. 139–167.
- Scheman, Naomi. 1997. "Queering the Center by Centering the Queer: Reflections on Transsexuals and Secular Jews." In *Feminists Rethink the Self*. Edited by Diana Tietjens Meyers. Boulder, CO: Westview Press, pp. 124–162.
- Scheman, Naomi. 2017. "On Mattering." In *Facts and Values: The Ethics and Metaphysics of Normativity*. Edited by Giancarlo Marchetti and Sarin Marchetti. New York: Routledge, pp. 119–136.
- Scott, James C. 1990. *Domination and the Arts of Resistance*. New Haven, CT: Yale University Press.
- Seo, Sarah A. 2019. *Policing the Open Road: How Cars Transformed American Freedom*. Cambridge, MA: Harvard University Press.
- Sewell, William, Jr. 2005. "The Concept(s) of Culture." In *Practicing History: New Directions in Historical Writing after the Linguistic Turn*. Edited by Gabrielle M. Spiegel. New York: Routledge.
- Shelby, Tommie. 2003. "Ideology, Racism, and Critical Social Theory." *The Philosophical Forum* 34(2): 153–188.
- Shelby, Tommie. 2014. "Racism, Moralism, and Social Criticism." *DuBois Review* 11(1): 57–74.
- Smith, Adam. 1776/1977. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edited by Edwin Cannan. Chicago, IL: University of Chicago Press.
- Solnit, Rebecca. 2016. *Hope in the Dark: Untold Histories, Wild Possibilities*. 2nd edition. Chicago, IL: Haymarket Books. Kindle Edition. <https://www.haymarketbooks.org/books/791-hope-in-the-dark>.
- Stahl, Titus. 2017. "Immanent Critique and Particular Moral Experience." *Critical Horizons*, <https://doi.org/10.1080/14409917.2017.1376939>.
- Taylor, Keeanga-Yamahtta. 2017. *How We Get Free: Black Feminism and the Combahee River Collective*. Chicago, IL: Haymarket Books.
- Toole, Briana. 2019. "From Standpoint Epistemology to Epistemic Oppression." *Hypatia* 34(4): 598–618.

- Tremain, Shelley. 2001. "On the Government of Disability." *Social Theory and Practice* 27(4): 617–636.
- Tuana, Nancy. 2017. "Feminist Epistemology: The Subject of Knowledge." In *The Routledge Handbook of Epistemic Injustice*. Edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr. New York: Routledge, pp. 125–138.
- Walker, Margaret Urban. 2007. *Moral Understandings*. New York: Oxford University Press.
- Walton, G. M. and S. T. Brady. 2017. "The Many Questions of Belonging." In *Handbook of Competence and Motivation (2nd edition): Theory and Application*. Edited by A. Elliot, C. Dweck, and D. Yeager. New York: Guilford Press, pp. 272–293.
- Wright, Erik Olin. 2010. *Envisioning Real Utopias*. New York: Verso.
- Young, Iris. 1990. *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press.
- Zawidzki, Tadeusz. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.

3

The Limited Use View of the Duty to Save

Helen Frowe

1. Introduction

This chapter defends the *Limited Use View* of the duty to save. The Limited Use View holds that the duty to save is a duty to treat oneself, and perhaps one's resources, as a means of preventing harm to others. But the duty to treat oneself as a means for the sake of others is limited; one need not treat oneself as a means when doing so is either very costly or conflicts with one's more stringent duties to others. By the same token, one's claim to be saved is limited. One has no claim that others treat themselves as means for one's sake when doing so is either very costly for them or conflicts with their more stringent duties to others.

In Section 2, I develop the Limited Use View of our duties to save. In Section 3, I consider the saving of special others. In Section 4, I explore the moral significance of the distinction between failing to save and depriving others of the means of saving themselves. I consider the implications of this distinction for the Limited Use View.

I then turn to a possible rival to the Limited Use View, which holds that our duties to rescue are limited by an agent-relative prerogative to weight our own interests, and the interest of special others, more heavily than the interests of others (Quong, 2009: 516; Fabre, 2012: 22). I argue, in Section 5, that this view faces a dilemma. To explain the limit on the duty to rescue, it must permit agents to grant very great weight to their own interests. And yet a permission to grant their own interests such weight leaves agents free to fail to prevent very serious harm when saving will cause them to incur even very moderate costs.

Several writers argue that since we have agent-relative prerogatives to fail to save, we must also have agent-relative prerogatives to do harm (Quong, 2009, 2016; Davis, 1984; Bazargan-Forward, in progress; Fabre, 2012: 61). According to these writers, the extra weight that we may attach to our own interests permits us to inflict ordinarily disproportionate harm on others to avoid harm to ourselves. Call this the *Broad Scope View* of agent-relative prerogatives. Along similar lines, Seth Lazar (2013) argues that we can have agent-relative duties to cause ordinarily disproportionate harm, since we are sometimes obliged to assign extra weight to the interests of certain special others.

I grant the conditional claim of the Broad Scope View. If agent-relative prerogatives bear on our duties to rescue, then they should also bear on our duties not to harm. Given this, the converse should also hold: if prerogatives do not bear on our duties not to harm, then we have reason to doubt that they bear on our duties to rescue. In Section 6, I reject several arguments that purport to show that agent-relative prerogatives, or associative duties, permit us to inflict ordinarily disproportionate harm for our own sake, or for the sake of special others. Section 7 argues that cases in which one may harm another person rather than harm oneself in order to save a third party also fail to vindicate the Broad Scope View. The Limited Use View can explain these cases without resorting to agent-relative prerogatives. Section 8 concludes.

2. The Limited Use View

Non-consequentialist moral theories typically hold that we have duties to rescue others from harm, but that these duties are limited. The limit is grounded in the cost to the prospective rescuer (and, of course, in the cost that the rescue might impose on others). According to these theories, one may ordinarily let a harm befall someone else in order not only to avoid a greater or equal harm to oneself, but also to avoid a somewhat lesser harm to oneself.¹ Consider *Self-Help Sharks*:

¹ I assume throughout that all the prospective victims and rescuers are innocent in the relevant sense—that is, nobody is culpable, morally responsible, or negligent with respect to

Self-Help Sharks: Aziz is being chased by Small Shark, who will bite off one of Aziz's legs if she catches him. Ben is being chased by Big Shark, who will bite off both of Ben's legs if she catches him. Aziz can either climb aboard a nearby boat, saving himself, or help Ben climb aboard, saving Ben.

It seems plausible that Aziz may save himself—that is, Aziz may deliberately fail to prevent a greater harm to Ben in order to prevent a smaller harm to himself.

And yet many non-consequentialists also subscribe to the view that, ordinarily, we ought to minimize harm when rescuing.² Consider *Rescue Sharks*:

Rescue Sharks: Aziz is being chased by Small Shark, who will bite off one of Aziz's legs if she catches him. Ben is being chased by Big Shark, who will bite off both of Ben's legs if she catches him. At no cost to herself, Rescuer (who is a stranger to Aziz and Ben) can help either Aziz or Ben onto her boat, but she cannot save both.

Let's assume that since she can prevent serious harm at no cost to herself, Rescuer is under a duty to rescue in *Rescue Sharks*.³ Plausibly, she cannot discharge this duty by saving Aziz, since Ben will suffer a greater harm. One cannot satisfy the duty to rescue simply by preventing some harm to someone, even if one thereby prevents a serious harm that would ordinarily give rise to a duty to rescue.⁴ After all, the duty to rescue *just is* a duty to prevent harm. It's hard to see how one might satisfy this duty by deliberately preventing less harm when preventing more harm is no costlier. Rescuer is thus under a duty to rescue Ben, and Ben has a corresponding claim that Rescuer save him, since he will otherwise suffer the greater harm. Saving Aziz thus conflicts with Rescuer's duty to save

what threatens the prospective victims. Things might be different, of course, if the rescuer wrongly endangered the victim. I also assume that there are no prior agreements that bear on whom it is permissible to save. For discussion of such agreements, see (Frowe, 2019).

² Of course, consequentialists believe this too.

³ The discussion in this section draws on (Frowe, 2019).

⁴ There might be some mild harms that I need not prevent even if I can easily do so—that is, that don't trigger duties to rescue at all.

Ben, and Aziz therefore has no claim to make use of Rescuer for his own sake.⁵

If Aziz and Ben face threats of equal harm, Rescuer ought to toss a coin to decide who to save, thereby giving each a fair chance of rescue. Even if both are strangers to Rescuer, we have enough evidence about people's biases—implicit or otherwise—to want her to engage in a fair decision-making procedure if she can.

I suggest that the permissibility of Aziz's failing to preventing a greater harm befalling Ben, despite the impermissibility of Rescuer's doing so, is explained by the fact that we each have only limited claims that others make themselves (and their legitimate resources) available for our sake. This limit is grounded in respect for persons, which involves recognizing that individuals are not means to be used for the general good. It is widely accepted that treating others as a means, especially in ways that impose costs on them, is very hard to justify. It is this that explains the impermissibility of lethally pushing a person in front of a trolley to save five, despite the intuitive permissibility of diverting a trolley away from five to where it will kill one as a side-effect (Foot, 1967; Thomson, 1985). But to demand that a person rescue is to demand that she treat herself as a means for the benefit of others: for example, that Rescuer devote herself to pulling Ben out of the water. Our claims that others treat themselves as means in this way is limited by both the costs that the prospective rescuer would thereby incur and by the rescuer's more stringent duties. For example, if Rescuer can reach Ben only by lethally mowing down a third person, Cathy, with her boat, her duty not to cause harm to Cathy defeats Ben's claim to be saved.

The fact that each of us has only a limited claim that others treat themselves as a means for our sake explains, in turn, why each of us has

⁵ There are, of course, alternative views of what Rescuer ought to do, such as use a weighted lottery, perhaps giving Ben a two-out-of-three chance of being rescued. These arguments are more popularly deployed in cases in which differently sized groups face harms of similar magnitude, since it is in these cases that, as Taurek notes, the loss is no better or worse for any particular person (Taurek, 1977: 307). I lack space to explore these alternatives here. My interest is the largely uncontroversial view that Rescuer may not simply save Aziz, even though Aziz may simply save himself. Those who prefer alternative accounts of Rescuer's duties should be able to substitute, *mutatis mutandis*, their preferred view for my claims about minimizing harm.

only limited duties to rescue. When making herself available for rescue passes a certain degree of cost to a prospective rescuer, the imperilled person typically lacks a claim that the prospective rescuer save her.

The claim that our duties to rescue are cost-sensitive in this way, such that victims simply lack claims to be rescued when saving them is very or disproportionately costly, is widely endorsed (Firth and Quong, 2012: 695; Quong, 2015: 252; Oberman, 2015: 258). However, what usually gets overlooked is that the cost matters specifically because the rescuer will be treating herself as a means for the victim's sake—that is, what matters is not merely the fact that one will bear a cost, but also how that cost will be incurred.⁶ We typically have no claim that others make themselves useful for our sake when doing so will impose significant or disparate costs on them. Rescuer need not save Ben from the loss of his legs at the cost of her own leg, or save Ben from a broken wrist at the cost of breaking her own leg.

Of course, one might be required to rescue even at very significant cost if the alternative is allowing a very grave harm. Perhaps one must jump in front of a trolley to prevent a hundred or a thousand deaths. But the fact that requiring a person to make very costly use of herself for the sake of others is justified only when the stakes are so high reflects the significance that we attach to harmful using compared to, say, harming as a foreseen side-effect.

We might think that, contrary to what I suggested in Section 1, agent-relative prerogatives explain why one must bear only a limited cost in the course of making oneself useful to others. Rather than rivalling the agent-relative view, perhaps the proposed limit arises precisely because one may attach greater weight to one's own interests than to the interests of others. But the Limited Use View does not rely on augmenting one's interests. Rather, it proposes an agent-neutral limit on our duty to rescue. It holds not only that one is not obliged to make *oneself* useful to others past a certain cost, but also that *others* may not make one useful past that

⁶ For example, Jonathan Quong points to the fact that our claims to be rescued are limited by the cost to the rescuer to support the view that our claims not to be harmed obtain only when one can 'reasonably demand' that others fulfil them (Quong, 2016: 253). Yet, as I argue below, this does not follow, since refraining from harming someone does not involve making oneself useful to her.

cost.⁷ As I discuss in Section 6, this does not entail that one may not be forced to bear an equivalent cost as a side-effect of a harm-preventing action. But it does rule out forcing other people to serve as a means when the cost of their doing so is sufficiently high that they need not treat themselves as a means. In other words, one cannot have a lesser-evil justification for forcing someone to serve as a means. Forcing someone to serve as a means is justified only if they are under a duty to treat themselves as a means. The right not to be used cannot be overridden by the good one can thereby accomplish. Correlatively, on this view, there is no difference between the point at which one is obliged to jump in front of a trolley and the point at which one would be obliged to push someone else in front of a trolley. If one must prevent the harm, and one can either jump oneself or push someone else, one ought to toss a coin.

3. Special Others

We might think that, by rejecting the idea of agent-relative prerogatives, the Limited Use View cannot accommodate the widely held intuition that one may sometimes save not only oneself but also certain ‘special others’ from a lesser harm. Consider *Parental Shark*:

Parental Shark: Aziz is being chased by Small Shark, who will bite off one of Aziz’s legs if she catches him. Ben is being chased by Big Shark, who will bite off both of Ben’s legs if he catches him. Mother, Aziz’s parent, but a stranger to Ben, can save either Aziz or Ben at no physical cost to herself, but she cannot save both.

It seems permissible for Mother to save Aziz, even though saving Ben would be no physically costlier for her. We would also likely grant that one may prevent the lesser harm if the person facing the lesser harm is one’s spouse, parent, sibling, or close friend. The proponent of agent-relative prerogatives has a ready explanation of these intuitive permissions. Our prerogatives permit us to weight not only our own interests

⁷ Thanks to an anonymous referee for pushing me to elucidate this difference.

but also the interests of special others more heavily (Scheffler, 1982: 21; Hurka, 2004: 59; Quong, 2020).⁸

It is tempting, since each of these permissions pertains to a special other, to suppose that the permissions share the same ground. But there is reason to doubt this. Some permissions to save special others from the lesser harm are plausibly grounded in the cost to the rescuer of failing to save, just like the permission to save ourselves from the lesser harm. If Aziz is Rescuer's close friend, for example, we can imagine that failing to save him would be very costly for Rescuer. But, like the permission to save ourselves, this permission is not grounded in an agent-relative prerogative to augment the interests of special others. Rather, it is grounded in the impersonal fact of how costly minimizing harm would be for the rescuer.

Whilst this cost explains Rescuer's permission to save Aziz if Aziz is her friend, it does not plausibly ground a duty for her to do so. After all, one can opt to bear very high costs for the sake of others—Rescuer could simply decide to bear the cost of seeing Aziz suffer, and perhaps the resultant damage to their friendship, in order to prevent the greater harm to Ben. Saving Ben is thus permissible but not required.

In contrast, Mother is plausibly required to save Aziz rather than Ben. This duty cannot be grounded in the cost to Mother of failing to save Aziz. Permissions grounded in costs to oneself can be waived, and Mother cannot waive her duty to save Aziz. Nor would an indifferent parent be required to minimize harm rather than save their own child, on the grounds that failing to save their own child is not costly to them. Parents plausibly owe it to their children to save them from a lesser harm even if they do not care about them. A duty to save one's child from a lesser harm must be grounded in some other fact arising from the relationship between parents and children. For example, Rivka Weinberg argues that parental obligations are grounded in parents' responsibility for exposing their child to the risks of life (Weinberg,

⁸ Precisely why we may favour special others in this way is contested. But it is generally agreed by proponents of agent-relative prerogatives that one may augment the interests of at least some special others.

2015).⁹ In general, responsibility for someone's being imperilled can alter one's duties to rescue. If Rescuer has shoved Aziz, but not Ben, into the water, Rescuer will have caused the loss of Aziz's leg if she does not save him. Rescuer's more stringent duty not to cause harm to Aziz defeats Ben's claim to be saved from even a somewhat greater harm. If something like Weinberg's account is correct, parents' responsibility for imposing risks of harm upon their child can similarly explain why Mother has a more stringent duty to save Aziz, if Aziz is her child, than she does to save Ben.¹⁰

Such duties are agent-relative in the sense that only those who are appropriately connected to Aziz's being imperilled are obliged to save him, since only those people are under a more stringent duty that defeats Ben's claim to be saved. But this is simply an instance of the general principle that one can, through one's actions or role, incur particular obligations. It does not require weighting one's own interests, or those of special others, more heavily than other people's, which is the form of agent-relativity that the Limited Use View rejects. Nor do these duties obtain only in the context of special relationships. If Rescuer has shoved Aziz in the lake, she ought to save him even if they are strangers.

I leave it open here precisely when we have duties, rather than permissions, to prevent the lesser harm.¹¹ Where there are such duties, these limit the claims that others have to make use of us, as per the Limited Use View. Ben lacks a claim that Mother make herself useful to him when Mother has a duty to save Aziz, just as Ben lacks a claim that Rescuer make herself useful to him if she can reach him only by mowing down Cathy.

⁹ Weinberg argues that these duties persist throughout a child's life. I remain neutral on this issue here. Even if one does not have a duty to save one's adult children from lesser harm, the cost of failing to do so would usually ground a permission to do so.

¹⁰ The causal view of parental obligations might support a similar explanation of a duty to prevent a lesser harm. See e.g. (Lindemann Nelson, 1991).

¹¹ We might think that we have duties to save our spouses from the lesser harm, for example, insofar as we make a commitment to prefer them above others. On the conditions for valid agreements to prevent the lesser harm, see (Frowe, 2019).

4. Depriving Others of Harm-Preventing Resources

Consider *Equal Canoe*:

Equal Canoe: Aziz is being chased by a shark who will bite off one of his legs. Ben is being chased by a shark who will bite off one of his legs. There is an abandoned one-person canoe nearby. Whoever reaches the canoe first will avoid his shark and suffer no harm. The other person will be bitten by his shark.

We might assume that if Aziz is permitted to save himself in *Self-Help Sharks*, where he climbs aboard the nearby boat rather than assisting Ben, then he must also be so permitted in *Equal Canoe*. After all, if Aziz may save himself from a lesser harm rather than Ben from a greater harm, he may surely save himself from a harm equal to that facing Ben rather than save Ben, as in *Equal Canoe*.

But I think this assumption is mistaken. *Self-Help Sharks* is a case of failing to save. By climbing aboard the boat, rather than assisting Ben, Aziz does not make Ben any worse off than Ben would have been in Aziz's absence. Aziz merely denies Ben the opportunity to benefit from his presence. Another way of putting this is that in *Self-Help Sharks*, Ben's being saved depends on Aziz's making himself useful to Ben. But Ben lacks a claim that Aziz make himself useful for his sake when this will be very costly for Aziz. Thus, if Aziz refuses to save Ben in *Self-Help Sharks*, Aziz does not thereby deprive Ben of a resource to which Ben has a claim.

But in *Equal Canoe*, whoever uses the canoe thereby deprives the other of a resource that he could have used to save himself, without needing anyone else to make themselves useful. This is more than failing to save the other person: if he takes the canoe, Aziz thereby prevents Ben from saving himself, and vice versa.

Aziz and Ben plausibly have equal claims to make use of the canoe. Since only one can make use of the canoe, their claims here amount to a claim to an equal chance of using the canoe. Just as Rescuer ought to toss a coin to determine whom to rescue when Aziz and Ben face equal

threats, Aziz and Ben should, ideally, toss a coin to see who may use the canoe. We should not simply grant that whoever reaches it first (and, perhaps, can fend off the other victim) is entitled to use the canoe. As Victor Tadros argues, we should avoid permitting the morally irrelevant fact of someone's greater physical prowess to determine who gets to use harm-preventing resources (Tadros, 2011: 208). The fact that Aziz is stronger than Ben, and can reach the canoe more quickly and then fight Ben off, does not give Aziz a permission to take the canoe.

Now consider *Unequal Canoe*:

Unequal Canoe: Aziz is being chased by Small Shark, who will bite off one of his legs. Ben is being chased by Big Shark, who will bite off both of his legs. There is an abandoned one-person canoe nearby. Whoever reaches the canoe first will avoid his shark and suffer no harm. The other person will be bitten by his shark.

The relationship between Aziz and Ben is the same as in *Equal Canoe*. Neither needs the other to make himself usefully available in order to avoid harm but, in taking the canoe, each will deprive the other of a resource that they could have used to save themselves. And yet in this case Ben will suffer a greater harm than Aziz. I argued in *Rescue Sharks* that facing the greater harm gives Ben a claim to be saved by Rescuer. We might think that changing the means of rescue from Rescuer to a canoe cannot effect any change in Ben's claims. If Ben has a claim to be saved by Rescuer in virtue of facing the greater harm, he also has a claim to save himself using the canoe.

But this is too quick. It matters that in *Rescue Sharks*, Aziz and Ben can avoid harm only via another agent, Rescuer. Since Rescuer ought to minimize harm, she has a duty to save Ben. Aziz has no claim that Rescuer make herself useful to him when doing so conflicts with her duty to prevent a greater harm by saving Ben. But in *Unequal Canoe*, Aziz's avoiding harm does not depend on another agent, on whose assistance he has no claim. He needs only to make use of himself and the canoe. In this respect, *Unequal Canoe* is akin to *Self-Help Sharks*, in which Aziz may climb aboard the boat rather than assisting Ben, even though Ben will suffer a greater harm.

Since in *Self-Help Sharks* Aziz does not, by saving himself, deprive Ben of a resource to which Ben has an equal claim, Aziz need not even toss a coin before saving himself. He may simply climb aboard the boat. But taking the canoe in *Unequal Canoe* does deprive Ben of a resource to which Ben also has a claim. Thus, whilst Aziz need not simply cede the canoe to Ben, he may not simply take it either. Rather, as in *Equal Canoe*, Aziz and Ben should toss a coin to see who gets to use the canoe. Ben's facing the greater harm does not entitle him to simply take the canoe. By depriving Aziz of a fair chance to use this harm-preventing resource, Ben would wrongly force Aziz to bear a cost for his sake that Aziz is not required to bear rather than save himself.¹²

What should we say about *Speedy Shark*?

Speedy Shark: Aziz and Ben are being chased by a shark. The shark will bite only the first person it reaches; the other will make it to shore. Aziz is tall, and so the shark will bite off only one of his legs. Ben is short, and so the shark, who is hungry, will bite off both of his legs. Aziz is a faster swimmer than Ben. He can keep ahead of Ben, making it safely to shore.

It seems permissible for Aziz to swim faster than Ben. But, again, this isn't a straightforward case of failing to save: rather, Aziz requires that Ben be bitten in order to avoid being bitten himself. *Speedy Shark* thus involves Aziz's making harmful use of Ben, rather than merely failing to rescue him. If Aziz outswims Ben, he does not require that Ben treat himself as a means. But he nonetheless forces Ben to become a means from which Aziz benefits.

However, this instance of harmful using is plausibly justified in light of the fact that Aziz and Ben are symmetrically situated with respect to harmfully using each other—that is, Aziz can escape only if Ben is bitten, and vice versa. When there is no way to avoid someone's being harmfully used, such using lacks its typical moral significance. Allowing the faster swimmer to simply take the canoe in *Equal Canoe* or *Unequal Canoe* is objectionable because it allows that swimmer to unfairly obtain a

¹² I consider the moral significance of ownership of resources in Frowe, 'Moral Constraints on Preventing the Lesser Harm' (unpublished ms.). I do not address that question here.

resource (the canoe) to which each threatened person has an equal claim. But in *Speedy Shark*, Aziz is not using his own resources (his swimming ability) to deprive Ben of a resource to which Ben has an equal claim. The only resource of which either will deprive the other in *Speedy Shark* is his own body. Given the cost involved, neither has a duty to make his body usefully available to the other. Thus, Aziz need not toss a coin to see whether he may swim faster than Ben.

To be clear: my claim is not that, provided that one is not making use of a person, or that someone's being harmfully used is unavoidable, one is permitted to cause harm in order to save oneself. It is, for example, plausibly impermissible for Aziz to defend his life against an attacker, Villain, if his defensive action will kill innocent Ben as a side-effect. It is also impermissible for Aziz to try to impede Ben's attempt to outswim the shark, for example by throwing bits of driftwood at Ben. Both harming and failing to save are subject to proportionality constraints, and the proportionality constraint on harming is more stringent than that on failing to save, given the moral worseness of doing harm compared to allowing harm. One must be securing considerably more good than one inflicts in order to justify harming a non-labile person, even when one does not make use of that person. The exception in *Speedy Shark* arises only because useful harming is both reciprocal and unavoidable.

5. Failing to Save as an Agent-Relative Prerogative

That we have only limited duties to treat ourselves as a means for the sake of others does the explanatory work that some accounts of the duty to rescue assign to an agent-relative prerogative. This prerogative supposedly allows each of us to weight our own interests, and the interests of special others, more heavily than the interests of strangers (Quong, 2009: 516–517; Quong, 2016: 817; Bazargan-Forward, 2018: 672–673; Fabre, 2012: 21, 57; Hurka, 2004: 59; Scheffler, 1982: 20).¹³ According to Jonathan Quong, this additional weight is “the only explanation” of

¹³ There is disagreement about how special relationships have to be to generate agent-relative prerogatives. I take no stance on that debate here.

why, for example, one may fail to save a child's life at the cost of becoming a paraplegic, even though the death of the child is the worse outcome, impartially considered (Quong, 2009: 517). Each person has "a powerful agent-relative permission to avoid sacrificing or significantly risking their own life for the sake of others" (Quong, 2009: 516–517).¹⁴ Seth Lazar advances a similar view about associative duties, arguing that we are morally required to confer additional weight on the interests of those to whom we stand in special relationships (Lazar, 2013: 21). In this and the following sections, I raise two objections to the agent-relative view of the limits on the duty to rescue. The first concerns the scope of the duty to rescue. The second concerns the scope of agent-relative prerogatives.

5.1. Augmenting Interests and the Scope of the Duty to Rescue

The agent-relative approach gets the intuitively right result in a range of cases. It allows, for example, that Aziz may save his one leg rather than both Ben's legs. But it nonetheless struggles to reconcile two plausible, widely held views about the scope of the duty to rescue. The first is that Aziz may fail to save not only Ben's legs but also Ben's life—indeed, several people's lives—in order to avoid the loss of his own leg. The second is that Aziz may not fail to save Ben's legs to avoid, say, the breaking of his own arm.

With respect to the first claim, imagine that Aziz can stop a runaway trolley from killing ten people only by placing his leg on the tracks, where it will be amputated by the trolley. We do not typically demand that people save at such great cost to themselves even when many lives are at stake. But to explain this on the agent-relative view, we must grant that Aziz may treat the loss of his one leg as being as bad as, or perhaps worse than, the death of ten people. In other words, securing this plausible claim requires that Aziz may attribute very significantly more weight to his own interests.

¹⁴ See also (Davis, 1984); (Bazargan-Forward, 2018: 671); (Fabre, 2012: 61).

Some (artificial) numbers help to illustrate the point. Say that, impartially considered, the loss of a leg has a disvalue of 10, and that death has a disvalue of 20. If Aziz may fail to save the ten, this means that he may fail to prevent a harm with a total disvalue of 200, in order to avoid a harm of 10 to himself. This means that harms to Aziz count for twenty times as much as harms to other people. And then consider that, plausibly, Aziz would not be required to suffer the loss of his leg to save even more than ten lives. Imagine that the trolley will kill twenty people unless Aziz places his leg on the tracks. I still doubt that he must place his leg on the tracks. If so, and if this is to be explained by his agent-relative prerogative, the extent to which Aziz may augment his interests must be greater still: harms to Aziz must count for at least forty times as much as harms to other people.

Now consider what this means for the second plausible claim. Imagine that Aziz can prevent the loss of Ben's two legs at the cost of breaking his own arm. He is, *prima facie*, under a duty to save Ben at this cost. But if Aziz may attribute to harms to himself the degree of extra weight suggested above, it doesn't look as if Aziz can be under a duty to save after all. Say that a broken arm has a disvalue of 3, and the loss of two legs has a disvalue of 15. Given the augmenting of Aziz's interests (they count for at least forty times other people's interests), his broken arm will turn out to be a great deal worse, from Aziz's perspective, than the loss of both Ben's legs. Hence, Aziz may refrain from saving Ben.

The dilemma, then, is that augmented interests can justify a plausible limit on the duty to save only if agents may augment their own interests by a considerable degree, such that the loss of one leg turns out to be worse than ten or twenty deaths. But if agents may augment their interests by such a degree, they may do so for any cost they will incur in a rescue. And this renders the duty to rescue implausibly undemanding. Even comparatively moderate costs, so inflated, will outweigh serious harms to others. This problem generalizes to other ways of augmenting agents' interests, such as allowing one to give one's interests some additional fixed weight. The fixed weight must be very significant in order to explain the limit on the duty to rescue, but such significant weight cannot ground a plausibly robust duty to rescue.

The underlying problem is that the concept of augmenting interests is at odds with the structure of the duty to rescue, which depends only

partly on comparing the harms at stake. A *prima facie* duty to rescue obtains when one could prevent a non-trivial harm to someone else.¹⁵ Whether one is in fact under duty then depends on a comparison of the harm facing the victim with the cost of rescuing her. But this is true only insofar as the cost of rescuing falls below a certain threshold—that is, below the cost that we can be expected to bear in the course of making ourselves useful to others. Once that threshold is reached, further harms to others do not weigh against the cost to the rescuer. The threshold itself is not generated by a comparison of the harms at stake. This is why Aziz is not required to rescue ten or twenty people at the cost of his leg: not because he may augment his own interests until they outweigh the other harms at stake, but rather because the threshold is insensitive to the other harms at stake. If the loss of his leg exceeds the limit Aziz must bear for the sake of saving others, further harms are typically irrelevant. On the augmenting view, in contrast, there is no cost threshold as such. Rather, the cost one can be required to bear is always settled by a comparison with the harm one can prevent—it's just that one's own interests are inflated to limit the duty's demandingness.

I say that harms beyond the threshold are *typically* irrelevant in order to leave room for a 'catastrophe clause', where one can have a duty to rescue (or force others to rescue) at very high cost in order to avoid a moral disaster (Nozick, 1971: 30(fn.)). As Saba Bazargan-Forward puts it, "we do not place infinite importance on our status as inviolable beings. We are willing to accept a regime that permits sacrificing an individual when doing so is necessary to avert a catastrophic harm because we do not think that any single person's status is so important as to warrant protection under those circumstances" (Bazargan-Forward, 2018: 690). As I suggested above, one might be required to jump in front of a trolley to save a thousand people. But there is nonetheless considerable space between the limit on our ordinary duties to rescue and the point at which a harm is sufficiently catastrophic to compel us to rescue beyond that limit. Within this space, our permission to refrain from saving is justified

¹⁵ As in note 4, we plausibly lack duties to prevent trivial harms to other. But one can deny this whilst still accepting the broader picture that I paint here.

not by comparisons with the harms at stake, but rather by the fact that saving breaches the cost threshold.

Contrast this with the agent-relative view, which must hold that one may fail to prevent the harms that fall into this space only if they are outweighed by the rescuer's augmented interests (recall Quong's insistence that this augmenting is the only possible explanation of why one may fail to prevent more serious harms to others to avoid a lesser cost to oneself). If this space includes harm up to the level of a genuine moral catastrophe, this view is deeply implausible, for the reasons just given. Moreover, as I argued above, the limit on our duty to rescue is not agent-relative. If Aziz is not required to make himself useful to Ben at a certain cost, others are not permitted to make him so useful either.

The more plausible account of our duties to rescue, then, seems to be that one may sometimes refrain from saving *even though* refraining does not produce the better outcome—considered either impersonally or from one's own perspective.¹⁶ It's true that some failures to save might *also* produce the better outcome, impartially considered. If Aziz can save Ben from a broken finger only by suffering a broken arm, failing to save is the better outcome. But this is not what explains the permission to refrain from saving. That is explained by the fact that Ben has no claim that Aziz makes himself useful to him at such cost, in order that Ben may avoid a smaller cost.

Denying that agents may augment harms to themselves is compatible with the claim that harms can be morally weighted for the purposes of proportionality calculations. We might, for example, believe that harms to a culpable or responsible attacker count for less than harms to her innocent victim, or to a bystander. But this moral weighting reflects the idea that, for example, the attacker's moral responsibility for an unjust threat decreases the moral significance of harms to her. It is much less plausible that such weighting can take place with respect to harms to non-labile people, such that harms to oneself or special others count for more in proportionality calculations, simply in virtue of being harms to oneself or a special other, compared to harms to other innocent people.

¹⁶ As above, there may be some amount of harm that Aziz is required to prevent even at the cost of losing his leg.

6. The Scope of Agent-Relative Prerogatives

According to the Limited Use View, the fact that one lacks a claim to be saved by an agent does not imply that one might also lack a claim not to be harmed by that agent. Consider *Defensive Shark*:

Defensive Shark: Aziz is being chased by Small Shark, who will bite off one of Aziz's legs if she catches him. Ben is on a nearby raft. Aziz can use a grenade to kill Small Shark, but the blast will also blow off both of Ben's legs.

The costs facing each party are the same as in *Self-Help Sharks*. In *Self-Help Sharks*, the cost that Aziz would bear in making himself useful to Ben defeats Ben's claim to be saved by Aziz. But if we require Aziz to refrain from throwing the grenade because doing so will harm Ben, we do not thereby require Aziz to make himself (or his resources) useful to Ben. Given this, the cost that will befall Aziz if he refrains from harming Ben lacks the purchase it has in rescue cases. Note that this is true even if we somewhat increase the cost to Aziz to reflect the moral significance of doing, rather than allowing, harm. Aziz may not throw the grenade even if he will otherwise lose both his legs, and the grenade will cause the loss of only one of Ben's legs.

Of course, one may sometimes harm others, both for one's own sake and for the sake of others. Such harming can be justified as the lesser evil. But lesser-evil justifications obtain only when harming secures substantially more good than the harm inflicted. As above, it is plausibly permissible to kill one as a side-effect of saving five lives, but not fewer than five. Importantly, lesser-evil justifications override, rather than defeat, agents' claims not to be harmed, and these claims continue to make demands upon us. For example, a person who is justifiably harmed on lesser-evil grounds is entitled to compensation. In contrast, we need not compensate people whom we justifiably fail to save.

The historically orthodox view of prerogatives supports this asymmetry, holding that while prerogatives bear on the permissibility of failing to save, they do not bear on the permissibility of harming (Scheffler, 2002; McMahan, 1997; Pogge, 2010). But it is doubtful that

a prerogative-based view can support this distinction. If an agent-relative prerogative is a permission to weight some people's interests more heavily than others, it's unclear why that permission would be limited to failing to save other people, and not extend to harming other people (Kagan, 1989; Kagan, 1984: 251). After all, one cares more in general about one's own interests, and those of one's nearest and dearest; one does not have a special concern that applies only in cases in which one has a *prima facie* duty to aid others.

Of course, the moral worseness of causing, rather than allowing, harm makes it harder to justify harming an innocent person in order to avoid harm to oneself compared to failing to save an innocent person in order to avoid harm to oneself. But this should generate only a difference in degree regarding how much harm one may impose for one's own sake, compared to how much harm one may allow for one's own sake. It does not support the view that prerogatives are relevant to failing to aid, but irrelevant to harming. If one's interests count for more, this should be reflected in the degree of harm one may inflict on both liable and non-liable people for one's own sake, compared to that which one may inflict for the sake of strangers.

This observation was first presented as an objection to the idea that agents have a prerogative to weight their own interests more heavily. But some proponents of prerogatives have instead embraced this Broad Scope View, arguing that prerogatives do indeed justify both failing to save and harming (Quong, 2009, 2016; Davis, 1984; Hurka, 2004: 61; Bazargan-Forward, *in progress*; Fabre, 2012: 61). For example, Jonathan Quong argues that our agent-relative prerogatives can permit us to cause harm in the absence of either a lesser-evil or liability-based justification for doing so, providing an independent justification for harming even innocent people (Quong, 2016). Lazar argues that our associative duties alter lesser-evil calculations, since harms to special others count for more than harms to others (Lazar, 2013: 18).

I agree that if agent-relative prerogatives underpin our permissions to fail to save, then they must also bear on our permissions to do harm. Thus, by showing that prerogatives do not plausibly bear on our permissions to do harm, we thereby undermine the view that they bear on our permissions to fail to save. In what follows, I reject three arguments for the view that prerogatives can justify harming.

6.1. Prerogatives as Independent Justifications for Harming

Quong has recently proposed two arguments in defence of the Broad Scope View (Quong, 2016). The first defends the claim that one has an agent-relative permission to kill innocent threateners who will make use of one's body. The second defends the claim that one has an agent-relative permission to kill innocent threateners who will not make use of one's body.

The first argument draws on Judith Jarvis Thomson's well-known case, *Violinist*.

Violinist: A famous musician, Violinist, is dying from a kidney ailment. Victim alone has a particular blood type that can cure him, if Violinist is connected to her body in a way that enables him to use her circulatory system for nine months. The Society of Music Lovers kidnap Victim and Violinist while they are both unconscious, and connect his body to hers. (Thomson, 1971: 49)

Quong argues that it is permissible for Victim to unplug herself from Violinist, knowing that this will cause Violinist's death. And yet, Violinist, being unconscious, is not responsible for being connected to Victim, and therefore not liable to be killed. There is also no lesser-evil justification for killing Violinist, since, impersonally considered, Violinist's death is a worse harm than the harm facing Victim. The only possible explanation of the permissibility of Victim's unplugging herself is that she has an agent-relative prerogative to do so. Hence, agent-relative prerogatives can justify harming. Violinist has a right not to be killed by Victim, but this right is permissibly infringed by Victim.

An upshot of this argument is that whilst Victim may unplug herself, other people may not unplug her even if she is explicitly refusing her consent to being used by Violinist in this way.¹⁷ From an agent-neutral

¹⁷ Quong briefly considers the idea that, at best, special others, e.g. Victim's spouse, may be permitted to unplug Victim, but does not offer a full defence of that claim. Since I don't think the permission to unplug is agent-relative, I need not engage with that claim here. But for

perspective, Quong argues, nothing could justify bringing about Violinist's death to "save another from nine months in a hospital bed" (Quong, 2016: 823).

Quong observes, though, that this argument might generate defensive rights of only fairly limited scope. Violinist is a particular type of innocent threatener—namely, one who must make use of Victim's body to survive. Since being made use of is morally significant, we might think that this at least partly explains why Victim may kill Violinist. And yet Quong thinks that our prerogatives are broader in scope than this. We may also kill people who do not need to make use of us, as in *Defensive Trolley*:

Defensive Trolley: A runaway trolley is heading to where it will kill five people. Frank, a bystander, diverts the trolley to where Erica is trapped. Erica can prevent herself from being killed by the trolley by throwing a grenade that destroys the trolley and kills Frank as a side-effect.

Quong argues that Erica may kill Frank. If Frank is permitted to kill Erica in the course of saving the five, "how can it be impermissible for Erica to change the result so that it is Frank, rather than her, who dies so long as the five will be saved either way?" (Quong, 2016: 826). Yet, he says, Frank is not liable to be killed—because he acts with objective justification—and Erica has no lesser-evil justification for killing him. Only an agent-relative prerogative can explain the permissibility of Erica's throwing the grenade that will kill Frank.

6.1.1. Duties and Using

I think that Quong is correct about the permissibility of disconnecting in *Violinist*, but wrong about what explains that permissibility. I think he is incorrect about the permissibility of killing Frank in *Defensive Trolley*.

It's true that, as things stand in *Violinist*, if Victim does nothing, Violinist will be saved. In a standard rescue case, it is only if the rescuer acts that the victim is saved. But this difference does not mean that

Victim is not saving Violinist: clearly, she *is* saving him, for as long as she allows him to make use of her body. The relevant question is whether Violinist has a claim to be saved by Victim at the cost to Victim of having someone make use of her body for nine months. And it seems uncontroversial that Violinist has no such claim, because this is why it is uncontroversial that the members of the Society of Music Lovers act wrongly in forcibly connecting Victim to Violinist in the first place.¹⁸

We can also see that Violinist lacks a claim to be connected to Victim by thinking about versions of the case in which Violinist is not unconscious throughout. If he is conscious when he is kidnapped, he ought not to agree to the plan—indeed, he ought to do his best to thwart it. Victim is under no duty to save Violinist at the cost of nine months' use of her body, and thus Violinist should try to prevent Victim's being forced to save him. Note that Violinist should do this even foreseeing the lethal cost to himself. I may fail to save someone to avoid a lethal cost to myself; I may not force someone else to save me to avoid a lethal cost to myself when doing so imposes on her a cost that she is not required to bear.

If Violinist wakes up after he has been connected to Victim, he has a duty to disconnect himself if Victim asks him to do so, or if Victim is unconscious and cannot give or refuse consent. Violinist cannot simply presume consent: at best, we allow hypothetical consent with respect to actions that are intended to benefit the person whose consent is presumed.¹⁹ If Violinist cannot disconnect himself, he must ask someone else to do it. And Quong agrees with this: for him, Violinist's being unconscious is essential precisely because, if he were awake, he ought to disconnect himself. *Intentionally* remaining connected to Victim would be a violation of her right not to be used, rendering Violinist liable to defensive harm (Quong, 2016: 822).

But if one lacks a right to use a person's body intentionally, one cannot acquire such a right by using it unconsciously. If Victim has a right that a conscious Violinist disconnect himself, or have himself disconnected by

¹⁸ As Quong (2016) grants on p. 823.

¹⁹ Or, perhaps, to benefit someone with whom she has a special relationship, but is under no duty to rescue. For example, we might think that it is permissible to use an unconscious person's rare blood type to save her brother's life, even if she has no duty to save him, if we reasonably believe she would consent if she could.

others, she may disconnect him or have him disconnected by others irrespective of whether he is conscious. Note that none of this requires us to take a stance on the relationship between agency and violating rights, or the relationship between violating rights and liability. It requires only the very plausible thought that Violinist lacks a right to make use of Victim's body—which, as above, explains why the members of the Society of Music Lovers act wrongly in connecting Violinist to Victim. If Violinist had such a right, the Society would not act wrongly in connecting them. Given the extent of the wrongful usage—nine continuous months—it is plausibly proportionate to lethally prevent the use of Victim's body by Violinist. The fact that the wrongful actions of others mean that Victim must now stop Violinist from using her, rather than merely fail to save him, cannot make it impermissible for Victim to disconnect herself.²⁰ If she has no duty to connect herself to Violinist, Victim has no duty to remain connected to him when she is forcibly and wrongly connected by others.

Quong's view that it would be impermissible for a third party to disconnect Victim, even if she is conscious and explicitly refusing consent to be used, is also implausible. His claim that third parties are choosing whether to kill one person, or save another from nine months in a hospital bed, obscures the morally salient features of the case (Quong, 2016: 823). On that description, *Violinist* is akin to a case in which a doctor must decide whether to remove a life-support machine from a patient whose life it will save in order to give it to a patient for whom it will avoid the need for nine months' bed rest. It thus fails to capture what is at stake in *Violinist*, where a doctor is deciding whether one patient should be allowed, for nine months, to make non-consensual use of another patient's body to benefit himself. Patients are not pieces of equipment that doctors may distribute to where they will do the most

²⁰ The Limited Use View thus secures defensive permissions against at least non-labile threateners who will otherwise make costly use of one's body, since defence in such cases is justified by the limits on our duties to rescue. It does not follow from this that one has defensive permissions against non-labile threateners *only* when they will otherwise make use of one's body. Although I reject (in the next section, and elsewhere) Quong's view that one may defend oneself against objectively justified threateners, I remain neutral here on the question of defence against non-labile, objectively unjustified threateners (such as a toddler with a gun, or person in the grip of a schizophrenic episode).

good, or leave where others have distributed them in the hope of doing the most good. Once we recognize that this costly use of Victim is what is at stake in *Violinist*, it is deeply counter-intuitive that a doctor may not prevent the non-consensual use of Victim by Violinist.

To see this even more plainly, imagine that being connected to Violinist will wear out Victim's kidneys, so that at the end of the nine months Violinist will be restored to health and Victim will die. But Violinist is unconscious, so his killing of Victim is unintentional (call this case *Lethal Violinist*). If Quong is correct that there is no morally relevant difference between Victim and Violinist, it seems that the doctor must let Violinist kill Victim, rather than disconnect him. After all, in terms of the outcome, it's no better that Victim live than Violinist. And given the moral significance of the distinction between doing and allowing, Quong's view implies that the doctor ought to allow Violinist to kill Victim, rather than kill Violinist to save Victim.

But that cannot be right. To be sure, both Violinist and Victim are innocent, in that neither is morally responsible for an unjust threat to the other. But it surely matters, from an agent-neutral perspective, that Violinist is lethally and non-consensually making use of Victim's body. Anyone may and should prevent the wrongful, non-consensual use of Victim's body (and we know it *is* wrongful, since we know that the Society act wrongly in connecting her, and Violinist would act wrongly in failing to disconnect himself). The fact that Quong's agent-relative account cannot accommodate a permission for a doctor to disconnect Victim should make us deeply suspicious of the claim that agent-relativity explains Victim's permission to disconnect herself. We do not need to invoke agent-relative prerogatives to explain Victim's permission to disconnect, and such prerogatives cannot explain the very plausible view that others have duties to disconnect her.

6.1.2. Defence Against Non-Using Threateners

Quong argues that agent-relative prerogatives also support a permission to kill non-liaible people who will not otherwise make use of one's body, such as Frank in *Defensive Trolley*. Quong suggests that there is no reason for the cost of saving the five to fall on Erica rather than Frank. Erica may therefore switch the trolley towards Frank even though there is

no lesser-evil justification for doing so, and Frank is not liable to be killed. Notice that nothing in Quong's argument limits Erica's prerogative to killing Frank: rather, it licenses Erica's killing any bystander as a consequence of saving her own life. Neither Frank nor a bystander violate Erica's right not to be killed; neither Frank nor a bystander is liable to be killed to save Erica.²¹

Quong suggests that Frank and Erica are on a moral par (Quong, 2016: 827). But this seems false. Granted, both are morally innocent. But, as we saw in *Lethal Violinist*, this does not entail across-the-board moral parity. It is permissible to lethally disconnect Violinist to save Victim even though Violinist is morally innocent with respect to the threat he poses. Given this, merely pointing to the equal innocence of Frank and Erica will not show them to be on a moral par. Not all innocent people have equal rights to kill. Frank poses an objectively justified threat that he is under no duty not to pose.²² By using force against him, then, Erica is not forcing Frank to comply with a duty that he has not to harm her. She is simply killing one person to avoid letting herself die, even though the person who will otherwise kill her is objectively permitted to kill her.

If, as I have argued, Victim's permission to disconnect in *Violinist* does not rest on agent-relative prerogatives, Quong cannot rely on *Violinist* to support the relevance of prerogatives in *Defensive Trolley*. And he cannot point to the limits on our duty to rescue to support a permission to kill Frank. Erica is not being forced to make herself useful to either the five or Frank. As I argued in Section 2, it matters *how* a cost is incurred. Costs to a prospective rescuer undermine the right not to be saved because we have only limited duties to make ourselves useful to others. The mere fact that Erica will bear a lethal cost if she does not kill Frank does not

²¹ Quong's view particularly struggles to accommodate cases in which one has a lesser-evil justification for harming. Such justifications obtain only when refraining from harming is too costly to be demanded: it is permissible for Frank to kill Erica because the alternative (five deaths) is unreasonably costly. On Quong's view, this should entail that Erica lacks a right not to be killed, since Quong argues that both our right to be saved and our right not to be harmed depend on whether it is reasonable to demand that others fulfil those rights. It's not reasonable to demand that Frank refrain from killing Erica. And yet, Quong argues that Erica not only retains her right not to be killed, but may also kill Frank in self-defence. For more on this, see (Frowe, 2015).

²² I argue elsewhere that Frank is required to turn the trolley. If so, the idea that he could be an appropriate target of force in virtue of doing what morality requires looks even more implausible. See (Frowe, 2018).

undermine Frank's right not to be harmed when Frank has no need to make use of Erica. Frank's right not to be harmed could be overridden by a lesser-evil justification, if Erica would otherwise bear a much greater cost than she would impose on Frank. But since killing Frank is not the lesser evil, it is impermissible for Erica to kill Frank.

6.2. Agent-Relative Duties and Lesser-Evil Justifications for Harming

Seth Lazar argues that associative duties bear on the permissibility of harming because "our deep personal relationships can affect how much harm can be inflicted as the lesser evil" (Lazar, 2013: 18). Lesser-evil justifications are not straightforward consequentialist justifications: although they are sensitive to outcomes, they are also sensitive to deontological considerations (Frowe, 2018; Rodin, 2011). For example, as I suggested in Section 2, one is plausibly justified in collaterally killing one for the sake of saving five, but unjustified in usefully killing one for the sake of saving five.

Lesser-evil justifications are usually thought to weigh goods and harms impartially, providing objective justifications for harming on which anyone may act (McMahan, 2015: 9; 2014: 109; Frowe, 2018). Lazar rejects this view, arguing that an agent's duty to treat harms to special others as especially weighty can generate lesser-evil justifications for harming that are indexed to the agent, permitting (indeed, requiring) her to inflict what would ordinarily be disproportionate harm.

Lazar defends this view by way of the following three cases:

Meteor: A single meteor is heading towards the earth. If Alice does nothing, it will kill five innocent people. Alice can use a missile launcher to divert the meteor away from the five towards another innocent person, Clare.

Lazar claims that in *Meteor* Alice ought to save the five rather than refrain from killing Clare. This tells us that saving the five is at least as important as not killing Clare. Now consider:

Two Meteors: Two meteors are heading towards the earth. One is heading towards five innocent people. One is heading towards Alice's daughter, Bethany. Alice can divert only one meteor, which will then land harmlessly in a field. The other meteor will kill whoever is in its path.

Lazar claims that in *Two Meteors* Alice ought to save Bethany rather than save the five. This tells us that saving Bethany is at least as important as saving the five. But, of course, this importance must be explained by the fact that Bethany is Alice's daughter—a stranger, with no agent-relative interests at stake, would not even be permitted to save Bethany instead of five other people, let alone required to do so.

Finally, consider *Diversion*:

Diversion: A single meteor is heading towards Alice's daughter, Bethany. Alice can divert the meteor to where it will kill Clare.

(Lazar, 2013: 19–20)

Lazar claims that in *Diversion* Alice ought to lethally divert the meteor away from Bethany and towards Clare. After all, we've seen that saving Bethany is, from Alice's perspective, at least as important as saving the five. And Alice is required to kill Clare as a side-effect of saving the five. Alice must therefore be required to kill Clare as a side-effect of saving Bethany. Hence, agent-relative duties can justify killing.

Lazar recognizes that most non-consequentialist accounts of agent-relative interests hold that such interests cannot justify harming. But he argues that such accounts are overly sensitive to the distinction between doing and allowing, and insufficiently sensitive to different types of harming. He suggests that while agent-relative interests cannot justify intentional, opportunistic harming, they can justify foreseeable, eliminative harming. He argues that Alice's killing of Clare in *Diversion* is an instance of foreseeable eliminative harming.²³

²³ Lazar's understanding of eliminative harming is at odds with other accounts in the literature, most of which describe killing Clare as a foreseen side-effect, rather than an eliminative killing. This need not detain us here.

Lazar says that his contention is merely that agent-relative duties bear on the degree of harm that one may inflict as the lesser evil (Lazar, 2013: 18). But his argument needs more than this: it goes through only if the amount of harm that Alice may *fail to prevent* is also determined by what is the lesser evil. Only then could the fact that Alice ought to fail to save the five in order to save Bethany in *Two Meteors* show that her agent-relative duties bear on lesser-evil justifications. And this is the crucial move: once we grant that agent-relative duties bear on lesser-evil justifications, and we agree that lesser-evil justifications can justify harming, we have strong support for Lazar's conclusion that one may harm on the basis of agent-relative duties.

But, as we saw in Section 5, the scope of our duties to rescue cannot be plausibly explained by lesser-evil considerations. We are not justified in failing to save only if doing so produces the better outcome. Aziz may refrain from placing his leg on the trolley tracks even though ten or twenty deaths is clearly the worse outcome, compared to the loss of one leg. Similarly, Alice ought to fail to save the five, in order to save her daughter, even though five deaths is clearly the worse outcome compared to one death. And the rights of the five are defeated, not overridden, in this case. The five (or their beneficiaries) are not owed compensation for Alice's failure to save them, as they would ordinarily be if their rights were infringed on lesser-evil grounds. If lesser-evil considerations do not explain why Alice ought to fail to save the five in *Two Meteors*, we have no reason to grant Lazar's view that agent-relative duties bear on lesser-evil justifications. And thus we have no support for the crucial move that agent-relative duties justify harming.

Thinking of claims to be saved as limited claims to make use of others, as I have proposed, gives us a different analysis of Alice's permissions in the Meteor cases. Alice ought to kill Clare as a side-effect of saving the five in *Meteor* because she has a lesser-evil justification for doing so, grounded in the fact that the death of five people is substantially worse than the death of one person, impartially considered.²⁴ Clare's right is overridden, not defeated. Her beneficiaries are plausibly

²⁴ I defend the view that acting on lesser-evil justifications for harming is required in (Frowe, 2018).

owed compensation. And Alice ought to fail to save the five in *Two Meteors* because it is a case of failing to aid, in which the five need Alice to treat herself as a means for their sake if they are to survive. The five have no claim that Alice treat herself as a means when doing so is either very costly for her or conflicts with her more stringent duties, including her duty to save her child. Since their claims are defeated, the five (or their beneficiaries) have no claim to be compensated in light of Alice's justified failure to save them.

But Clare in *Diversion* does not need Alice to make herself useful for Clare's sake, if Clare is to survive. She merely needs Alice to refrain from killing her. Of course, Lazar believes that our duty to save our child is more stringent than our duty not to kill a stranger. But in the absence of transitivity from Alice's permission to fail to save to a permission to inflict harm, we have no argument for the view that the duty to save one's child is more stringent than the duty not to kill. And it is doubtful that Clare's claim not to be killed is defeated by Alice's duty to save her child, as the claims of the five to be saved are defeated in *Two Meteors*. As above, Clare's beneficiaries are owed compensation in *Meteor*, where Clare is killed to save five. Given this, her beneficiaries are surely owed compensation in *Diversion*, where Clare is killed to save one. If so, Clare's claim not to be killed is not defeated by Alice's duty to save her child in the way that a claim to be saved can be so defeated.

I argued, in Section 2, that we can have agent-relative duties to prevent the lesser harm, and that these duties can be explained by, for example, an agent's being responsible for a victim's being in danger. If, for example, Rescuer has shoved Aziz into the shark-infested waters, she will have caused the loss of Aziz's leg if she does not save him. Given this, it is permissible for Rescuer to save Aziz's one leg at the cost of failing to prevent the loss of Ben's two legs. But this does not entail that Rescuer may cause the loss of Ben's two legs in the course of saving Aziz. Nor would she be permitted to cause an equivalent harm to Ben—the loss of one leg—in the course of saving Aziz's leg. Aziz's claim not to be harmed is no more stringent than Ben's claim not to be harmed. Similarly, even if we grant that parents should save their children at the cost of allowing greater harm to others, this does not entail that parents may cause harm to others to prevent a lesser or equal harm to their child. Rather, they

may cause only harm that is justified as the lesser evil, impersonally considered.

7. Choosing Between Deaths

Before closing, it is worth addressing a challenge that proponents of the Broad Scope View might level against the Limited Use View. Consider *Busy Trolley*:

Busy Trolley: A runaway trolley is lethally heading towards five people. Bystander can lethally divert the trolley towards herself, or towards innocent Workman, who is trapped on a side-track.

Most people believe that Bystander need not divert the trolley towards herself, but that it is permissible for her to divert it towards Workman.²⁵ If this is correct, it looks as if agent-relative prerogatives might justify harming after all. This would explain why Bystander may kill Workman rather than kill herself.

But the Limited Use View can also explain the permissibility of Bystander's killing Workman. In order to be saved, the five need Bystander to make use of herself for their sake: they need her to divert the trolley. Since making use of herself in this way will impose a lethal cost on Bystander, she may refrain from using herself for the sake of the five.²⁶ They have no claim to be saved when doing so imposes this cost on the prospective rescuer. But, as we noted in *Defensive Trolley* in the previous section, diverting towards Workman does not involve making use of Workman, or forcing Workman to make use of himself. Thus, even though the five have no claim to make use of Workman if doing so will force him to bear a lethal cost, they do have a claim that Bystander save them by diverting towards Workman, since Bystander does not

²⁵ The case is in Thomson (2008: 364), where she attributes the idea to Alexander Friedman. Thomson argues that since it would be supererogatory for Bystander to divert towards herself, she may not divert towards Workman. But few people seem persuaded by this. See, for example, (FitzPatrick, 2009); (Walen and Wasserman, 2012); (Tadros, 2011: 120).

²⁶ Tadros suggests this in (Tadros, 2011: 120).

thereby use Workman as a means of saving the five. The five are being rescued by Bystander, not by Workman.

It's morally significant, then, that Bystander's options are either harming Workman or making harmful use of herself. Contrast *Busy Trolley* with *Busy Trolley Two*:

Busy Trolley Two: A runaway trolley is lethally heading towards Victim. She can divert the trolley down a side-track to where it will kill five innocent people, divert it down a different side-track to where it will kill Engineer, or let herself be killed by the trolley.

Victim has a choice between letting herself die and killing either Engineer or the five. Killing the five to save herself is clearly impermissible. But so too is killing Engineer. Such a killing cannot be justified as a side-effect of saving the five, since the five are not presently threatened. And Victim is not saving either the five or Engineer by refraining from diverting towards them. This is not a case of imposing a harm that is justified on lesser-evil grounds, but rather a case in which one must either kill a non-labile person or persons or let oneself die. Since there is no lesser-evil justification that can overcome Engineer's right not to be killed, Victim must refrain from diverting the trolley.

Similar reasoning applies in *Busy Trolley Three*:

Busy Trolley Three: A runaway trolley is lethally heading towards five people. Victim can divert it away from the five down a side-track, on which Victim herself is standing. The side-track has a branch, down which Victim can again divert the trolley before it hits her. Controller is standing on this branch. Whoever is hit by the trolley will be killed.²⁷

One might think that since Victim is permitted to divert the trolley away from the five towards Workman in *Busy Trolley*, she may divert away from the five and then divert again towards Controller in *Busy Trolley*

²⁷ Thanks to Krister Bykvist for this case.

Three. But this would be a mistake. In *Busy Trolley Three*, killing Controller cannot be justified as a side-effect of saving the five: they are saved when the trolley is first diverted. If Victim subsequently diverts down the branch towards Controller, she is saving only herself, which is disproportionate and hence impermissible. In *Busy Trolley Three*, then, Victim may refrain from saving the five, since they have no claim that Victim make lethal use of herself for their sake, and she can permissibly save them only in a way that involves making lethal use of herself. But if she chooses to save the five, she may not then divert the trolley towards Controller. Killing Controller is not the lesser evil and, as a killing, cannot be justified on the grounds that Controller has only a limited claim that Victim make herself useful for his sake.

8. Conclusion

The Limited Use View conceives of our duties to rescue as limited duties to treat ourselves as a means for the sake of others. We have no duty to treat ourselves as a means when doing so is unreasonably costly for us, or conflicts with our other, more stringent duties to others. Correlatively, we each have only limited claims to be rescued.

That we have only limited duties to make ourselves useful for the sake of others is an upshot of the claim that individuals are not means to be used for securing the greater good. Allowing individuals the space to pursue their autonomously chosen projects requires that we limit the claims that others can make on individuals. Within this space, one need not minimize harm.

The Limited Use View offers a largely agent-neutral account of our duties to save. If an agent is not required to make use of herself for the sake of preventing harm to others, it is impermissible for others to make use of her to that same end, if doing so will impose at least as much cost upon her. The cost to a rescuer of failing to save certain special others gives us a similarly agent-neutral explanation of other permissions to prevent the lesser harm. The Limited Use View also grants that there can be agent-relative duties to prevent the lesser harm. But such duties are not explained by a prerogative to augment the interests of either the

rescuer or the victim. It is this type of agent-relativity that the Limited Use View rejects.²⁸

Cited Works

- Bazargan-Forward, S., 'Vesting Agent-Relative Permissions in a Proxy', *Law and Philosophy* (2018), 37: 671–695.
- Bazargan-Forward, S., *Complicity in War* (unpublished manuscript) (in progress).
- Davis, N.A., 'Abortion and Self-Defense', *Philosophy and Public Affairs* (1984), 13/3: 175–207.
- Fabre, C., *Cosmopolitan War* (Oxford: Oxford University Press, 2012).
- Firth, J.M., & Quong, J. 'Necessity, Moral Liability and Defensive Harm', *Law and Philosophy* (2012), 31: 673–701.
- FitzPatrick, W., 'Thomson's Turn on the Trolley', *Analysis* (2009), 69/4: 636–643.
- Foot, P., 'The Problem of Abortion and the Doctrine of Double Effect', *Oxford Review* (1967), 5: 5–15.
- Frowe, H., 'Claim Rights, Duties, and Lesser-Evil Justifications', *Proceedings of the Aristotelian Society: Supplementary Volume* (2015), 89: 267–285.
- Frowe, H., 'Why We're Required to Turn the Trolley: On Lesser-Evil Justifications for Harming', *The Philosophical Quarterly* (2018), 68/272: 460–480.
- Frowe, H., 'If You'll Be My Bodyguard: Agreements to Save and the Duty to Minimize Harm', *Ethics* (2019), 129/2: 204–229.
- Frowe, H., 'Moral Constraints on Preventing the Lesser Harm' (unpublished manuscript).

²⁸ Drafts of this chapter were presented at the Oxford Studies Workshop in Political Philosophy, Syracuse University; the Stockholm Centre for the Ethics of War and Peace's Philosophy, Ethics and Killing (PEAK) workshop; the Moral and Political Philosophy seminar at Helsinki University; the Philosophy Department Colloquium at the University of Reading; and the Philosophy Department Colloquium at Stellenbosch University. I am very grateful to participants at each of these events. Special thanks to Rosa Terlazzo, who served as my respondent at Syracuse, and to the editors and referees for OSPP. This research was funded by a Knut and Alice Wallenberg Fellowship (grant number 1521101).

- Hurka, T., 'Proportionality and the Morality of War', *Philosophy and Public Affairs* (2004), 33/1: 34–66.
- Kagan, S., 'Does Consequentialism Demand Too Much? Recent Work on the Limits of Obligation', *Philosophy and Public Affairs* (1984), 13/3: 239–254.
- Kagan, S., *The Limits of Morality* (New York: Oxford University Press, 1989): 19–24.
- Lazar, S., 'Associative Duties and the Ethics of Killing in War', *Journal of Practical Ethics* (2013), 1/1: 3–48.
- Lindemann Nelson, J., 'Parental Obligations and the Ethics of Surrogacy: A Causal Perspective', *Public Affairs Quarterly* (1991), 5: 49–61.
- McMahan, J., 'The Limits of National Partiality', in Robert McKim and Jeff McMahan (eds) *The Morality of Nationalism* (Oxford: Oxford University Press, 1997): 107–138.
- McMahan, J., 'Self-Defence Against Justified Threateners', in H. Frowe and G. Lang (eds) *How We Fight: Ethics in War* (Oxford: Oxford University Press, 2014): 104–137.
- McMahan, J., 'Proportionality and Necessity in *Jus in Bello*', in S. Lazar and H. Frowe (eds.) *Oxford Handbook of the Ethics of War* (New York: Oxford University Press, 2015).
- Oberman, K., 'The Myth of the Morally Optional War', *Philosophy and Public Affairs* (2015), 43/4: 255–286.
- Pogge, T., 'Cosmopolitanism: A Defense', *Critical Review of International Social and Political Philosophy* (2010) 5/3: 86–91.
- Quong, J., 'Killing in Self-Defense', *Ethics* (2009), 119/3: 507–537.
- Quong, J., 'Rights Against Harm', *Proceedings of the Aristotelian Society: Supplementary Volume* (2015), 89: 249–266.
- Quong, J., 'Agent-Relative Prerogatives to Do Harm', *Criminal Law and Philosophy* (2016), 10: 815–829.
- Quong, J., *The Morality of Defensive Force* (New York: Oxford University Press, 2020).
- Rodin, D., 'Justifying Harm', *Ethics* (2011) 112/1: 74–110.
- Scheffler, S., *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982).
- Scheffler, S., *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought* (Oxford: Oxford University Press, 2002).

- Tadros, V., *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011).
- Taurek, J., 'Should the Numbers Count?', *Philosophy and Public Affairs* (1977), 6/4: 293–316.
- Thomson, J.J., 'A Defense of Abortion', *Philosophy and Public Affairs* (1971), 1/1: 47–66.
- Thomson, J.J., 'The Trolley Problem', *Yale Law Journal* (1985), 94/6: 1395–1415.
- Thomson, J.J., 'Turning the Trolley', *Philosophy and Public Affairs* (2008), 36/4: 359–374.
- Walen, A., & Wasserman, D., 'Agents, Impartiality and the Priority of Claims Over Duties: Diagnosing Why Thomson Still Gets the Trolley Problem Wrong by Appeal to the "Mechanics of Claims"', *Journal of Moral Philosophy* (2012), 9: 545–571.
- Weinberg, R., *The Risk of a Lifetime* (New York: Oxford University Press, 2015).

4

Politics, Truth, and Respect

Jonathan Turner

1. Introduction

Worries about truth in politics tend to fall into two groups. Philosophical discussion has focused on an alleged association of authoritarianism and claims to truth.¹ Recent political events, on the other hand, have left many equally concerned about the authoritarian implications of disregard for the truth. In this chapter I want to address an aspect of this topic that has so far received little explicit attention: the implications of the state's relationship to truth for its attitude towards its citizens.

My argument arises from some comments made by David Estlund in a recent defence of political liberalism (2012: 253):

Rawls argued that we must try to devise a conception of justice that does not purport to be the truth about justice, since that is something about which reasonable views disagree. He called this substitute account a “political conception” of justice... The idea of “reasonable” points of view strikes some people as insulting to other (“unreasonable”) views, but it is important to keep in mind that one of the main positions being opposed by this doctrine is that viewpoints that are mistaken, no matter how reasonable they might be, may be ignored completely. The Rawlsian denies this, and is thus more tolerant and liberal in an important sense. The whole point of this Rawlsian view is to give moral standing to certain viewpoints even if they are mistaken. Compared with traditional political philosophy, it vastly expands the range of views that must be accommodated in the terms of political

¹ See Arendt (1977), Estlund (2008: 21–39).

justification. The idea that it is an exclusionary philosophy is, it seems to me, simply a confusion.

In this passage Estlund describes what Raz (1994a) has called the “epistemic abstinence” of political liberalism. By this I understand—as I take Estlund to understand—two related things: first, the refusal of political liberalism *qua* theory to claim that the tenets of its political conception are true, rather than merely reasonable, and second, the policy stance of a state arranged along the lines advocated by political liberalism of rejecting viewpoints on the ground that they are unreasonable rather than untrue. (I take these to be features of any “political liberal” view.) My target is the claim that political liberalism is *pro tanto* less exclusionary than any comprehensive liberal view in virtue of its epistemic abstinence.

Estlund suggests, rather than fully articulating, the political liberal argument. I think his comments are instructive for understanding a certain dispute between political and comprehensive liberalism, and I will continue to refer to them, but I do not want to imply that the political liberal position that I reconstruct and criticize here derives directly from him. The political liberal argument I take from his comments is along the following lines. All reasonable viewpoints deserve moral standing. (I will shortly suggest that one way to interpret this requirement is as an aspect of the requirement of respect for persons.) Comprehensive liberalism denies this standing to reasonable but (allegedly) mistaken viewpoints because it regards their being mistaken as sufficient reason to ignore them. The charge of exclusionariness is therefore more appropriately levelled at comprehensive liberalism than at political liberalism.

Here, briefly and imprecisely, is the response to this argument that I will spend the rest of this chapter filling out. It is true that the comprehensive liberal alternative ignores views that are mistaken (although, significantly, it doesn’t ignore the fact *that people hold those views*), but it does so on the basis that we are all interested in what really is just, and we are all concerned together in trying to find out what that is. None of us wants a false view to be accepted—because it is false. So if I am told that my view is not accepted because it is false, that is at least a justification that treats me with respect, as engaged in a joint project of

discovering true justice. Of course, I still disagree: I think my view is true. (Otherwise it wouldn't *be* my view, because beliefs aim at truth.) But there is nothing unusual or insulting or exclusionary about a disagreement as to the truth.² There is no problem, then, in the state's regarding people's views as false; it is how the state responds to the fact that its citizens hold these beliefs that is important: that is what will tell us whether or not it respects them. Treating a person with respect involves responding appropriately to the relevant facts about her, one of which may be that she holds a false view, rather than bracketing one of those facts, as epistemic abstinence seeks to do.

This chapter has four further sections. In (2) I present the case for understanding exclusion in terms of respect, and explore the idea of insult, which is also mentioned in Estlund's comments. In (3) I explain the crucial distinction between ignoring a false view and ignoring the fact that someone holds that view. In (4) I use this distinction to argue that it is not exclusionary to tell citizens that their view is not accepted because it is false. In (5) I conclude.

2. Exclusion, Insult, and Respect

What I am concerned with in this chapter is the *attitude* that political liberalism endorses towards citizens, not with the substantive policies it generates. I do not ask whether political liberalism is more "tolerant and liberal" in the sense that it is likely to allow a greater range of personal freedom, or sanction less state coercion, than the comprehensive

² I make no positive case here for appealing to truth in political justification. An anonymous reviewer points out that without such a positive case, the comprehensive liberal may "look exclusionary in the sense that he antagonizes for no good reason people who otherwise could sign on to the conception of justice." Joshua Cohen (2008: 14–15) makes a convincing case for thinking that the concept (but not a theory) of truth is indispensable in political discourse in virtue of its necessary attachment to other ideas—not least judgement and belief—that such discourse cannot do without. It also seems to me that once one understands the minimal but necessary commitment involved in using the concept of truth, one should also be more inclined to wonder what objectionable commitment is involved in labelling another citizen's view as untrue that could constitute good grounds for thinking the label disrespectful of that citizen. Recognition of the indispensability of the concept of truth tends, I would suggest, to go hand in hand with recognition of the banality—the *unthreateningness*, one might say—of its predication.

liberalism to which it is opposed. I am concerned, as I take Estlund at least in part to be, with the justificatory stance taken by the state and by citizens towards the policies that they advocate. I am concerned, that is, with the endorsement of the “public reason” approach to democratic deliberation and debate. This approach “requires that the moral or political rules that regulate our common life be, in some sense, justifiable or acceptable to all those persons over whom the rules purport to have authority” (Quong 2018). So in considering whether epistemic abstinence is exclusionary or non-exclusionary in itself, we have to assume that the policies advocated by political and comprehensive liberalisms are held constant. This is not, I think, an unrealistic constraint, given that across a range of issues political and comprehensive liberal views (think Rawlsians and Kantians³) are likely to advocate similar policies. I will therefore say nothing about the policy implications of political liberalism, except in section 3, where I note the important difference between “ignoring” false views and ignoring the fact that people hold those views. Exclusion in the sense with which I am concerned is to be understood as something like a communication that one is not a proper member of the political community, rather than the marginalization or censure of one’s practices or way of life.⁴

My focus, then, is on a narrow category of exclusion. I am happy to accept that the comprehensive liberal view is exclusionary in some respects, as long as I can show why the view that Estlund takes to be confused actually makes sense; that is, why it makes sense to regard the comprehensive liberal view as respecting citizens in precisely the way that the political liberal view claims it does not.

The exclusion I am concerned about is the exclusion of citizens, not of views. The exclusion of citizens does not follow from the exclusion of the actually or possibly held views, as I explain in due course. Note that Estlund’s suggestion is that the idea of “reasonable” points of view could be insulting *to other views*. But it is hard to see how a view can be insulted. Even if this possibility is coherent, surely what matters here is

³ Assuming you can accept that Kant’s political theory is a comprehensive liberalism. See Pogge (1997), Hodgson (2010) (against), Nussbaum (2003: 26), Rawls (1993: 99) (for).

⁴ Nor should exclusion from membership of the political community be confused with exclusion from democratic participation.

whether *people* are insulted, on account of their views being labelled unreasonable. If political liberalism does make views rather than people the primary object of exclusion and inclusion, then it could be seen as less exclusionary than comprehensive liberalism, since it accords “standing” to a wider range of views (all reasonable ones) than does comprehensive liberalism (only true ones). But I don’t think this is the kind of inclusiveness that those who accuse political liberalism of being an “exclusionary philosophy” have in mind. Let me explain.

The idea of labelling some points of view as reasonable and others unreasonable is insulting—goes the comprehensive liberal criticism—not to unreasonable views, but to those who hold them: to the allegedly unreasonable people. The political liberal misses the point if she seeks to parry the accusation of insult against the supposedly unreasonable by pointing to its granting justificatory relevance to reasonable but untrue views (as against comprehensive liberalism’s exclusion of all but true views). The complaint is that I am being told my view is unreasonable: that is what excludes me from the political community. It is no consolation for me to hear that all reasonable views have justificatory relevance, even if they are held by no one, because I hold a view that has been labelled unreasonable. If we are to assess the relative exclusionariness of political and comprehensive liberalisms, we must at least be talking about the same kind of exclusion. In the sense that matters to comprehensive liberals, political liberalism is in the business of exclusion where comprehensive liberalism is not, because it is crucial to the political liberal project to identify citizens—the unreasonable ones—who do not count as full members of the political community.

I will interpret exclusion in terms of the notion of respect. There are several reasons for this. First, Rawlsian liberal views in general draw heavily and explicitly on the Kantian tradition in moral and political philosophy, and make extensive use of the Kantian notion of respect for persons.⁵ The concerns articulated by political liberals and other anti-perfectionists about comprehensive and perfectionist liberal views are often put in terms of the latter’s failure to respect citizens who favour unorthodox conceptions of the good, or who subscribe to “comprehensive” views

⁵ See, for example, Rawls (1971: 251–257, 586).

that the state judges reasonable but false.⁶ We should distinguish, however, between three such concerns. The first is the suggestion that the state disrespects its citizens by making policy on the basis of a contested conception of the human good. That is the *anti-perfectionist concern*. The second is that the state disrespects its citizens by making policy on the basis of reasons that they cannot accept. That is the *broad public reason concern*.⁷ It is a concern about the justification of coercion in circumstances of reasonable disagreement.⁸ The third is that the state disrespects its citizens by rejecting their views on the ground that they are untrue, but does not do so if it rejects them on the ground that they are unreasonable. That is the *narrow public reason concern*. It is a concern not about when it is permissible for the state to coerce citizens, but rather about what kind of message it is appropriate for the state to communicate to its citizens, if it is to treat them with respect.

My argument respects the anti-perfectionist concern: it is compatible with comprehensive anti-perfectionism. I believe it tells against the narrow public reason concern, but not necessarily against the broad concern. I argue that the state does not communicate disrespect to its citizens when it rejects their views as untrue. One might accept this point, but still insist that it is disrespectful to citizens if the state justifies coercive policy on the basis of reasons that citizens cannot reasonably accept. This is, in effect, to argue that it is inclusion of *views* that is really important. Why should the state not make policy on the basis of reasonably rejectable views? Political liberals answer this question in different ways, and a rebuttal of these arguments would be a full argument

⁶ Nussbaum (2011: 18). Larmore (1999: 608) regards political liberalism as being grounded in a principle of respect for persons but seeks to avoid "the peculiarities of the Kantian conception of respect". Other political liberals explicitly eschew any grounding in respect for persons: see Quong (2010: 230).

⁷ The anti-perfectionist concern may (and usually does) *derive from* the broad public reason concern but need not. Conceptions of the human good are simply one case of the kind of reasonable disagreement that drives the broad public reason concern. But the anti-perfectionist concern could be defended on other grounds; for instance, on the ground that the state shows its citizens equal concern and respect only by refraining from considering some kinds of preference in its policy decisions. See Dworkin (2000).

⁸ This is summed up by Rawls' (1993: 136) "liberal principle of legitimacy": "[O]ur exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason."

against political liberalism's distinctive account of political justification. I do not offer one here. But insofar as that account is motivated by a deeper concern for respect—by the thought that it is disrespectful to citizens to make policy on the basis of views they could reasonably reject, even if no one does in fact reject them—I explain in section 3 why I think my picture of the relationship between state and citizen is a viable alternative to the picture offered by political liberals motivated by this broader concern, and I hope that this alternative picture will undermine the appeal of that concern. But I do not argue directly against it. Insofar as the broad public reason concern is motivated by other values than respect—such as civic friendship, for example⁹—my arguments do not purport to address it.

The second reason to interpret exclusion in terms of respect is the way in which Estlund phrases his objection. The Rawlsian view, he says, gives “moral standing to certain viewpoints even if they are mistaken”. On one popular and, to my mind, compelling account, moral standing is precisely what a certain kind of respect is about. Stephen Darwall (1977: 38) distinguishes *appraisal respect* from *recognition respect*. Appraisal respect “consists in an attitude of positive appraisal of [a] person either as a person or as engaged in some particular pursuit.” It is an attitude that one deserves in virtue of one's actions or characteristics. Recognition respect, by contrast, is a “disposition to weigh appropriately in one's deliberations some feature of [a] thing...and to act accordingly” (Darwall 1977: 38). “[T]o respect something in this sense,” says Darwall (2004: 49), “is to *give it standing in one's relations to it*.” Recognition respect of this kind is the respect that is owed to all persons simply in virtue of their being moral persons.¹⁰

Finally, the idea of respect affords an intuitively appealing interpretation of the kind of exclusion that is involved here. We are not talking about physical exclusion, or about disabling someone from enjoying society's benefits, or barring her from political or any other forms of participation. We are talking about a more abstract sense in which

⁹ Leland (2019).

¹⁰ Darwall seems now to have changed, or at least supplemented, his view on respect for persons: see Darwall (2009). I don't agree with the new view but nor do I propose to argue against it.

comprehensive liberalism might be regarded as an exclusionary philosophy. This is a sense that is closely related to the idea of *insult*. Insult is a *communicative* form of disrespect. I fail to respect someone, in the recognition sense, by failing to take proper account of her moral standing in my actions towards her. So I fail to respect someone if I push her out of the way in my effort to board a train before its doors close. But in doing so I do not insult her. I treat her *as if* she were a mere object, but not in a way that suggests that I think she *is* a mere object. It is the calling into question of the existence or significance of some fact, rather than simply the failure to take proper account of it, that distinguishes insult from mere disrespect. Compare a couple of examples from the *Cambridge Dictionary* entry on insult:

The steelworkers' leader rejected the two percent pay rise saying it was an insult to the profession.¹¹

In offering such a meagre pay rise the employer not only fails to *take proper account of* (say) the steelworkers' past hard work, or the stagnation in their wages over the last several years, or the importance of the domestic steel industry in the current economic climate—but actually communicates that it *does not recognize* the significance of these things, or outright *denies* them.¹² Perhaps if the proposed rise had been 5 per cent, it would still have fallen short of the union's demand for 8 per cent, and would therefore, on their account, still have failed to take proper account of the steelworkers' labour, but it would not have been an insult because one could believe that the employer did not seek to deny or deprecate any of the facts cited by the workers in support of their proposed rise; they just calculated their implications differently. To disrespect someone, then, is to fail to take full account of relevant facts about her—her moral standing or intellect, the steelworkers' recent outstanding industry, and so on—but insult is more than this: it is to

¹¹ *Cambridge Dictionary* online, available at: <<https://dictionary.cambridge.org/dictionary/english/insult>> (accessed 20.02.18).

¹² My account of insult here has some affinities with Harry Frankfurt's (1997) account of respect.

impugn that moral standing, intellect, or industry. Hence the second example and its dictionary “translation”:

The instructions are so easy they are an insult to your intelligence.
(=They seem to suggest you are not clever if you need to use them.)¹³

As the translation makes clear, what is insulting as such is not the failure to take proper account of a relevant characteristic, but the suggestion that its very presence is in doubt. The worry about political liberalism that Estlund’s remarks suggest is that it is insulting in just this way, because it *impugns* the moral standing of those who hold unreasonable views. But, runs the reply, even if this charge can be substantiated, the comprehensive liberal fares worse in this regard. For her theory fails altogether to recognize even reasonable views if they are false.¹⁴

My reply to this is that comprehensive liberalism is guilty of no disrespect because it *does* recognize the only relevant fact that is in play when the view in question is false, namely the fact *that certain citizens hold this view*.¹⁵ To explain this, in the next section I take up the distinction between ignoring a false view and ignoring the fact that someone holds a view that happens to be false. First, a slight diversion. It should be acknowledged that the question whether political liberalism is exclusionary in the sense just sketched must be to some degree empirical. This is because the question whether the state’s communication to its citizens disrespects or insults them is partly determined by their actual reactions to that communication. A statement might reasonably be regarded as revealing something about the speaker’s—in this case, the state’s—attitude even if mistakenly so. And if such an interpretation is reasonable, such a statement might sensibly be regarded as

¹³ *Cambridge Dictionary* online, available at: <<https://dictionary.cambridge.org/dictionary/english/insult>> (accessed 20.02.18).

¹⁴ As an anonymous reviewer reminds me, Estlund frames his argument in terms of the exclusion of viewpoints rather than of those who hold them. But, as I have explained above, it seems to me that it is exclusion of citizens that matters.

¹⁵ Or might come to hold such a view. Even if no citizen currently holds a certain false view, the fact that they might come to do so is relevant for the state’s deliberations and pronouncements. Does this grant standing to false views? I don’t think so. It is still the relation between citizen and view that counts, rather than anything about the view considered independently.

disrespectful or insulting in spite of the purity of intention on the part of the speaker. As regards the kind of communication under consideration here, it seems to me not at all obvious that, as a matter of empirical fact, disregarding a person's view as unreasonable would engender less rancour, resentment, alienation, and so on than denouncing it as untrue. Consider, by way of illustration, the possible range of reactions to Martha Nussbaum's two imagined presidential speeches. (I discuss this example in section 4.) She acknowledges that Allen Wood finds the comprehensive liberal speech more appealing. I suspect that many others would too, especially if the comprehensive view was more fairly represented. At any rate, the considered responses of real citizens to the political liberal and comprehensive liberal stances seem to me of some importance to the question which kind of state better respects its citizens.¹⁶

3. Ignoring a False View versus Ignoring the Fact That Someone Holds a View (That Happens to Be False)

Estlund says that, as far as the comprehensive liberal is concerned, "viewpoints that are mistaken, no matter how reasonable they might be, may be ignored completely". In a sense, this is true. A view is reasonable in the moralized sense used in Rawlsian political liberalism if it is compatible with a willingness to "propose fair terms of cooperation" and to recognize the burdens of judgement (Rawls 1993: 48–66). It may be reasonable in this sense, and also in the sense of being a view that a person of normal powers of reasoning might arrive at after an appropriate¹⁷ degree of reflection, but for all that it may be false. And false views, or false beliefs, are of no value in rational or moral deliberation. In that sense, certainly, they may be ignored. What must be taken into account in any form of deliberation is the relevant facts. False beliefs

¹⁶ Wall (2014: 482) makes a similar point regarding the question whether the adherents of a religious way of life would "experience expressive subordination, even if they are aware of the fact that the ground for the state's tolerance and respectful attitude toward their way of life lies in an endorsement of a theory of value that they do not accept."

¹⁷ I mean "appropriate" in the sense relevant to epistemic justification rather than in any moral sense.

disclose no relevant facts, because *ex hypothesi* they disclose no facts at all; that is why they can be ignored. But there is a fact here that is relevant: that is the fact that someone holds this false view. And ignoring a false view (because it is false) is not the same thing as ignoring the fact that someone holds such a view.¹⁸

This distinction is a cousin of one made by Joseph Raz (1998: 27) and recently pressed by David Enoch (2013: 159–160). That is the distinction between the fact that *p* being a reason for state action, and the fact that I (in a position of state authority) believe that *p* not being a reason for anything. This distinction is supposed to defuse the charge that perfectionist liberalism gives illicit preference to the beliefs of state actors about the good over the beliefs of citizens in general. It gives no such preference, goes the counter-argument, because where I am in a position of authority, no significance is being attached as such to the fact that *p* is *my* view; the only thing that is relevant to my action is the fact (as I believe it to be) that *p*.¹⁹ In the Raz-Enoch counter-argument, what is important is that the fact that the state holds a certain view (the belief that *p*) does not have normative significance, whereas the truth of that view (the fact that *p*) does. Here things are the other way around. The (putative) fact that *p* has no significance, because there is no such fact²⁰ (although the fact that not-*p* certainly is significant!), whereas the fact that certain citizens *believe* that *p* does.

¹⁸ Jeremy Waldron (1999: 303–304) claims that treating our fellow citizens' *views* with respect requires that all possible options be up for grabs by majoritarian procedure in a democracy. But there is no sense to be attached to respect for false views *qua* views, and I doubt whether respect for *persons* involves allowing them to be harmed by terrible democratic decisions. There is still a question about how to take into account in our treatment of a certain group the fact that they believe (falsely) that *p*, and this may partly be a question of what respect for them requires (I comment on this below), but this is different from the question what respect for their false view requires.

¹⁹ What sometimes seems to be underplayed in this argument is that there is still a problem here about how the state comes to have the decisive say on what the facts are that are to be acted on. This, I take it, is sometimes the political liberal's actual concern. (This problem of privileged judgement is one that is more prominent in explicitly Kantian versions of political liberalism. See, e.g., Pallikkathayil (2010), Sinclair (2018)—although neither Pallikkathayil nor Sinclair are specifically defending political liberalism.) But this is quite different from the suggestion that the state's view is to be acted on because it is the state's view.

²⁰ For those suspicious about the ontology of facts: it is not significant that *p*, because *p* is false.

The point is this. A person's belief that *p* is false does not in itself carry many implications as to how we should treat her, where that includes how we should treat her *on account of her believing that p*.²¹ More specifically, that she falsely believes that *p* does not in itself tell us much about the kind of policies that might be appropriate in the area that the belief concerns. The policy implications of the fact that citizens hold a certain false view will depend on the circumstances, but here are some important possibilities:²²

(1) The government might take some action *in addition* to pursuing policy P in order to recognize, in addition to the fact that *p*, on which P is based, the fact that some people believe that not-*p*.

For example, the state might make some public pronouncement that recognizes that the policy pursued is reasonably controversial and will not find universal support. Such pronouncements are common in medical law cases. Where judges have to decide whether conjoined twins may be separated with the certain result that one will die,²³ or whether a child in a persistent vegetative state may be allowed to die,²⁴ they rightly feel compelled to acknowledge the rational force of the arguments on the other side, and the dissentients' sincere moral commitment to their views. If courts did not qualify their judgments in this way, the state could justifiably be accused of excluding some people by failing to take adequate account of their belief that (for instance) it was immoral to withdraw life support from such a child.²⁵

(2) The government might adjust its policy to take account of the fact that some people disagree with its normative basis.

Suppose that the current regime for organ donation after death is an opt-in one. Organs for transplant are taken only from those who have

²¹ This line of thought is suggested by Raz's remarks (1994b: 104–106).

²² I assume in all these cases that the beliefs concerned are integral to the relevant citizens' "comprehensive doctrines". I also assume that the disagreements concerned may plausibly be regarded, in at least some cases, as reasonable.

²³ *In Re A (Children) (Conjoined Twins: Surgical Separation)* [2001] Fam 147.

²⁴ <https://www.theguardian.com/uk-news/2018/feb/20/parents-lose-legal-fight-keep-liverpool-toddler-life-support> (accessed 20.02.18).

²⁵ I take it that a policy is pursued here insofar as the judgment of the court sets a precedent as to how patients in relevantly similar circumstances may be treated. (In this example I use 'government' in the broadest sense, so as to encompass the judiciary and legislature as well as the executive.)

explicitly asked that they be. Following a review, a government reaches the conclusion that the morally best regime is a compulsory one. But many citizens have deep-seated objections to such a regime. In recognition of this, but also of the need to increase the number of organs available, perhaps the government ought instead to institute an opt-out system. But the justification of the decision not to move to a compulsory regime is not that it is illegitimate to make policy on the basis of a contentious view about the ethics of organ donation, but that it would be, all things considered, wrong to impose a policy of compulsory donation on people who have deep-seated moral objections to it.²⁶ Where p is subject to reasonable disagreement, the state respects those who believe that not- p , not by refraining from acting on the belief that p , but by acting on this belief *and also* on the knowledge that some citizens believe that not- p .

(3) Where the government pursues a policy on the basis of the fact that p , it need not prevent people from acting on the belief that not- p .

Many aspects of even people's comprehensive views have little bearing on justice, so the state need not care what they believe or how they act on their beliefs. Consider astrological beliefs.²⁷ If some astrological beliefs were true, it might be sensible for a government to invest in predictive techniques based on astrology. Governments do not so invest because (perhaps among other reasons) they take such beliefs to be false. If challenged as to its neglect of the potential of astrological prediction, a government might reasonably respond by pointing out the falsity of these doctrines. But it need not also be committed to any intervention to change people's beliefs or to prevent action on the basis of them.²⁸

(4) Where the government pursues a policy on the basis of the fact that p , it may try to persuade citizens who believe not- p of their error.

²⁶ That is not to say that *any* policy imposed in such circumstances would be wrongly imposed. That view allows legitimacy a disproportionate weight as against justice. See Stemplowska and Swift (2018).

²⁷ For this example see Nussbaum (2011: 25–29).

²⁸ The state may of course regulate the provision of astrological services such as fortune-telling, but that is because of their potential harm rather than because of their basis in false beliefs.

Any government of the United States that sought to enact significant gun control measures would encounter significant, perhaps insuperable, opposition. But let's assume that the issue is sufficiently important that the government should not compromise as in (2). How should it respect those who believe that the right to bear arms trumps the increased risk of death by shooting to American citizens? It should try to persuade them that they are wrong. At least in the case of those who are open to rational engagement with the arguments, the state respects its dissenting citizens by acknowledging their dissent and addressing it head-on. It includes them in the political community and tries to prevent them from feeling alienated from the laws to which they are subject.²⁹

Note that only one of these options (2) involves the government actually changing its policy. In this case it is a necessary, but not a sufficient, condition of the change in policy that the objection be a reasonable one. Cases (1) and (4) describe ways in which the government might communicate with those who hold contrary views. In these cases it seems to me that the government bears at least some burden of explaining its policy or explicitly recognizing the dissenting view even to those whose objections are unreasonable, albeit that they might turn out to be unreceptive, either because they do not understand (their unreasonableness is epistemic) or they do not care (their unreasonableness is moral).³⁰

²⁹ Victor Tadros (2016: 135–157) regards the danger of alienation as a source of motivation for the broad public reason concern. This is relevant to Jonathan Quong's recent response to what I have called the Raz-Enoch argument. Quong (2016: 143) argues that "Persons have an important interest in not being required to subordinate their judgments about fundamental matters provided the judgments are compatible with the protection of the same important interests of others", and that "antiperfectionist political institutions" offer more equal protection of this interest than perfectionist ones. In his response Raz (2016: 180–181) doubts whether persons have any such interest, noting that what people really care about is the content of the rules to which they are subject, rather than the fact that those are rules that they do not endorse. I think this is right, but it glosses over an important point: that even if dissenters from the prevailing rules "acknowledge that if they are mistaken, and if the rules are justified . . . then it is a good thing that they are subject to these rules even though they are of the (mistaken) opinion that the rules are bad or wrong", those dissenters will still feel alienated from their political institutions and civic life. Having to subordinate one's judgement on fundamental matters is bad insofar as it works against my interest in not feeling so alienated. The setback to this interest must then be weighed against the gains to other dimensions of well-being that a perfectionist state is not debarred on principle from pursuing. Quong believes that this trade-off too speaks in favour of anti-perfectionism. For a contrasting view see the last section of Enoch (2013).

³⁰ Clearly there is more to say about the state's obligations, but I cannot offer a comprehensive account here. Note, however, that I use "reasonable" in a non-technical sense, so that my

The broad public reason concern indicates one way of taking into account the fact that some people reasonably believe that p when not- p . Those who choose this option worry that the state will disrespect those citizens if it makes policy on the basis that not- p (not on the basis of its belief that not- p !), and therefore seek to remove certain reasons altogether from the public arena. This way the state will avoid making policy on the basis of reasons some of its citizens reasonably reject. But this is not the only way of dealing with the fact that some people reasonably believe things that are untrue. The more appropriate way for the state to take into account the fact that some people reasonably but falsely believe that p is to act in a way that is sensitive *both* to the fact that p and to the fact that some people reasonably believe that not- p . I have suggested some ways in which it might do this.

The caricature of comprehensive liberalism would have it that if citizens hold a false view, p , all the state has to take account of in this vicinity is the fact that not- p . But this is not so: it must also take account of the fact that some of its citizens believe that p . And how it does so will go a long way to determining whether it respects those citizens. Political liberals, and anti-perfectionists in general, typically generate traction for their view by appealing to the importance for citizens of their deeply held beliefs. It is disrespectful, they think, to fail to take adequate account of those deeply held beliefs. But while respecting someone may involve trying to understand her point of view, recognizing that it is (in some cases) important to her, and possibly making allowances that enable her to act on those beliefs, it is not so obvious that such respect involves refraining from acting on a belief that she doesn't—and, given her deeply held convictions, is highly unlikely to³¹—share.

So although the exclusionary scope of the comprehensive view is great in the sense that only true views are valuable in themselves, in a more

being unreasonable does not imply—as it does for the political liberal—that I am excluded from the constituency of justification of government policy.

³¹ I don't say "can't share" because it's difficult to see how best to identify what beliefs an individual could, given a "respectable amount" of reasoning (Gaus 2011: 250), or cognitive psychotherapy (Brandt 1979: 11), or the elimination of all logical inconsistency, and so on, come to hold. Given that people *have* undergone radical shifts in worldview, and that some views *are* true and others false, a category of beliefs that people can't share seems to me artificial and unhelpful.

important sense the exclusionary scope of that view depends on what it has to say about the implications for our treatment of people that they hold certain views.

4. Why Telling Citizens That Their View Is Not Accepted Because It Is False Is Not Exclusionary

Let's consider in the light of the foregoing distinction how the idea of recognition respect helps us to understand what might be meant by the suggestion that political liberalism gives moral standing to viewpoints even if they are mistaken.³² First, we should ask: is it really the viewpoint to which we grant moral standing, or the person who holds it? I think things make more sense if it is the person. Suppose it is the view itself which should be given moral standing. That is to say that the view should be respected. How do we give standing to a false view? The question is equivalent to asking how we are to weigh some feature of the view appropriately in our deliberations, and act accordingly. The most pertinent feature of the view is that it is false. The most appropriate weight to attach to a false view, it seems to me, is zero; and I act accordingly by ignoring the view. The view, *qua* view, simply has no moral relevance. Another way to put it is this: if I accord recognition respect to an object, what I do is take appropriate consideration of certain facts about the object. In this case there are no relevant facts about the view.³³ But there is a relevant fact about the person whose view it is, namely the fact that she holds the view. This is why it makes more sense to think about according moral standing to persons. (And this is how objections to perfectionist and comprehensive liberal theories are usually put.³⁴)

³² I'm not sure whether Estlund regards "viewpoint" as synonymous with "view". A viewpoint might be regarded as something in between a view and the person who holds the view, something more like "outlook". But Estlund's talk of "mistaken" viewpoints suggests he means no more than "view", and I talk simply of "views" from now on.

³³ If we put things more precisely, no recognition is due to a true view either, *qua* view, since recognition is due to the facts that the view reflects rather than any facts about the view itself. I do not mean to suggest that there is any non-pragmatic way to decide which objects facts are "about". Whether there are or not has no bearing on the argument in the text.

³⁴ Nussbaum (2011: 33) takes it to be important that respect is for persons, not views, but I am not convinced that she observes this consistently. See, for example, (2011: 28): "Why not let

The state's obligation to accord its citizens respect is an obligation to give them their proper standing. Now it is easy to see how the state might be thought to *act* with disrespect towards its citizens by forcing upon them the sort of life that they cannot endorse from within their own conception of the good—it does so because (the argument goes) it fails to take proper account of their capacity to respond to reasons.³⁵ I hope I have also shown in section 3 that a state might disrespect its citizens by failing to take appropriate account in its actions of the fact that some of its citizens hold certain false views. But it is not so easy to see how the state might *communicate* disrespect to its citizens—insult them—simply by claiming or implying that their views are untrue rather than unreasonable. In fact, as I will argue in this section, I recognize the moral standing of others precisely by acknowledging their participation in a common inquiry into the truth about morality.³⁶

The assumption we make in respectfully rejecting citizens' views as false is that those citizens are honestly seeking after the truth about justice, or whatever value is in question, just as we are. Political liberalism, of course, does not regard it as a proper function of political philosophy to reason towards the truth on political matters,³⁷ but this does not hold of everyday citizens in real societies (that is, most people insofar as they are not political liberal philosophers); nor does it hold of the citizens imagined by political liberalism as adhering to a fully or partially comprehensive view from which they are able to derive support for an overlapping consensus around a political conception of justice. Political liberalism itself, in other words, accepts that what people believe, they believe to be the truth. As truth-seekers, none of us is

[those who believe in astrology], and their beliefs, alone?" It may be this conflation that leads Nussbaum to the odd view that a public official's statement that astrology is not well grounded is disrespectful to the believer in astrology.

³⁵ I doubt this. In fact, I take Wall's view (2014: 479) that "Respect for persons must take account of their capacity to respond to reasons, both epistemic and practical. For the state to view citizens as inevitably tied to their comprehensive commitments, irrespective of the reasons they have for these commitments, is to fail to respect them as rational agents." But I don't pursue the anti-perfectionist concern here.

³⁶ As Waldron (1999: 305) recognizes: the "alternative to the self-interest model" of politics, he observes, is one in which citizens are "motivated in their disagreement not by what's in it for them but by a desire to get it right." The view I am defending does not depend on seeing politics as involving "high-minded citizens converging on truth".

³⁷ On this aspect of political liberalism, see especially Quong (2010).

infallible, or claims to be. I understand that I may sometimes get things wrong. Although I may in the instant case be adamant that I have not done so, I have some experience—as we all do—of realizing that what I previously thought was true was, in fact, mistaken. I therefore have no trouble in understanding the justification that is given for the rejection of my view, namely that on this occasion I have got it wrong. In fact, we can go further than this. This justification—“your view is wrong”—is one that I *must* accept as reasonable and intelligible, because it is precisely my own belief that the state has got it wrong that leads me to continue to oppose its view. I can understand why the other side wants to reject my view precisely because I want to reject theirs for the same reason.

At this point a familiar objection may be raised. I am supposed to content myself with the thought I may make mistakes like any other person, but the state is entitled to prefer its own view to mine on every occasion. Isn't the state in this story supposing itself to be infallible? It is not. The state is supposing itself to have the authority to decide which view to act on, and in the case of some states that supposition may itself be unjustified. But the arguments I make here are compatible with a variety of stories about how the state might ground a valid claim to such authority. There are some things the state must do: it must acknowledge its own belief-forming mechanisms to be fallible: it must be open to counter-evidence; it must be appropriately restrained when acting on its beliefs; and it must show itself to be suitably chastened when shown to have been wrong.³⁸ (These precepts apply most conspicuously to representatives of the state such as government ministers.) But no agent can do better than act on the basis of what she believes to be true, and the state is no exception.

I suggested in the previous section that the exclusionary scope of comprehensive liberalism depends on what it has to say about the implications for the state's treatment of people that they hold certain (false) views. But perhaps it might be thought that for the state publicly to reject a view as false rather than unreasonable is *already* a form of unequal

³⁸ As Mill (1989: 21) notes: “for while every one well knows himself to be fallible, few think it necessary to take any precautions against their own fallibility, or admit the supposition that any opinion, of which they feel very certain, may be one of the examples of the error to which they acknowledge themselves to be liable.”

treatment, before any further policy ramifications are considered. Martha Nussbaum apparently takes this view. She suggests (2011: 22) that comprehensive doctrines

are so deeply a part of people's search for the meaning of life that public governmental denigration of those doctrines puts those people at a disadvantage, suggesting that they are less worthy than other citizens, and, in effect, not treating them as fully equal ends in themselves.

But it is hard to see the warrant for such an inference.³⁹ That someone believes a false view may put her at a disadvantage—for example, a person may compromise her life expectancy through her religious objection to blood transfusion or other medical interventions—but the state's public endorsement of the contrary view does not do so, in the absence of discriminatory legislative, executive, or judicial treatment on the ground of her view.⁴⁰ And the state's rejection of her view gives it no ground for any such discrimination. Nussbaum simply states, rather than arguing for, the claim that denial of the truth of one's beliefs necessarily involves unequal treatment.

The claim about worthiness is similarly unsupported. Democratic societies have by and large come to accept⁴¹ by now that certain characteristics, most notably sex, gender, and ethnicity, are not grounds for regarding a person as less worthy of moral consideration or legal

³⁹ As long as we put aside Nussbaum's tendentious use of "denigration." As I understand her, Nussbaum at least takes the denial of the truth of a doctrine to amount to a denigration both of that doctrine ("What Rawls wishes to rule out is that the state would make statements (or incorporate principles) denigrating one religion or doctrine and preferring another, so long as the doctrine in question is 'reasonable'" (2011: 29)) and of its adherents ("For a public official in a leading role to say 'X's doctrine is not as well grounded as Y's' is, inevitably, to denigrate X" (2011: 33)). On the basis of this usage, I take the "public governmental denigration" quoted in the text to refer to no more than an authorized assertion by a public official of the falsity of a doctrine.

⁴⁰ The possibility of disadvantage to citizens on account of their believing a view the state rejects—perhaps through their being shunned as fools—raises the prospect of an *instrumental* reason for epistemic abstinence: governments might adopt no official view on comprehensive doctrines, not out of respect for their adherents, but for fear of the adverse consequences that might befall them.

⁴¹ At least in (constitutional) theory. I am very far from suggesting that discrimination on the basis of these criteria has been eradicated.

recognition. Religious and other “comprehensive” belief is obviously another such characteristic. Mill, for instance, is not committed by reason of his comprehensive liberalism to any view on the moral worth of citizens beyond his utilitarian creed that each is to count for one and no more than one.⁴² As for respecting people as ends in themselves, this Kantian injunction is about according recognition respect to people’s status as rational beings. The comprehensive liberal state respects a person’s capacity for reason in just the way that I have suggested: by occupying the same epistemic ground as she does on the question of her belief, taking her to be a sincere but fallible inquirer after the truth, rather than patronizing her by remaining neutral in a way that she does not.

Finally in this section, let me bolster the explanation given here of why rejecting views as false is not exclusionary by considering some related state attitudes that *could* be rightly denounced as exclusionary. Nussbaum imagines two presidential speeches addressed to constituencies of citizens who reject the state-endorsed doctrine of the equality of the sexes, the first by a comprehensive liberal president, the second from a political liberal. The comprehensive liberal denounces the view that women ought to be ethically subordinate to men as not only sexist, but its contrary as “self-evident”. No doubt it does insult people to tell them that in their search for the moral truth, not only have they gone wrong, but they have missed what ought to have been glaringly obvious (what is self-evident). There is enough disagreement about the truth of moral propositions, let alone their self-evidence. Any suggestion that a deeply contested moral proposition is self-evident is bound to be controversial. This makes the accusation of self-evidence a gratuitous one. The implication is that those who take the wrong view are not really being earnest in their search for the moral truth, or they are falling below what must be regarded as a respectable standard of success, or they are just being

⁴² Again, unless we take seriously the language of “denigration”; but comprehensive liberalism is not committed to the denigration of any view that political liberalism would regard as reasonable. It might be objected that the other cases are not analogous because, while one may have the (objectively) wrong beliefs, one cannot have the “wrong” sex, gender, or ethnicity. True: but the move from taking a person to believe what is false to regarding her as less worthy of moral consideration remains in need of justification.

obtain.⁴³ But there is no need for the comprehensive liberal to say more than that they have got it wrong.⁴⁴

Nussbaum (2011: 19) quotes Jacques Maritain (1961) in support of her political liberal conception of tolerance, but it is interesting that his words include reference to precisely the element in the comprehensive liberal view that I have suggested is crucial to its appeal as a respectful philosophy:

There is real and genuine tolerance only when a man is firmly and absolutely convinced of a truth, or of what he holds to be a truth, and when he at the same time recognizes the right of those who deny this truth to exist, and to contradict him, and to speak their own mind, not because they are free from truth but because they seek truth in their own way, and because he respects in them human nature and human dignity.

In other words, it is precisely our recognition that we are each in parallel earnestly engaged in a search for truth about justice that enables us respectfully to reject each other's views as false. In doing so we are doing no more than holding each other up to the standard to which we hold ourselves.⁴⁵

5. Conclusion

Comprehensive liberals believe that the state should be responsive to the facts. Political liberals deny this. They think certain facts should be

⁴³ One might actually *think* any of these things, but it is precisely because the state has a responsibility to show respect to its citizens that the president should refrain from saying so.

⁴⁴ Nussbaum's comprehensive liberal president also denounces her sexist citizens' views as "unAmerican". This seems to me a peculiarly American (and to British—and perhaps other European—ears, rather odd) insult, and not worth dwelling on, but certainly just as gratuitous as the self-evidence charge.

⁴⁵ A further thought on Maritain: he famously responded to astonishment at the possibility of agreement on fundamental rights among the members of the drafting committee for the UN Declaration of Human Rights by saying, "Yes, we agree about the rights but on condition no one asks us why." But the committee members did not substitute agreement for truth-seeking; rather, they agreed not to push their truth-seeking too far so as not to jeopardize their agreement. The worry is not about the commitment *to truth* as such, but about the *extent* of the commitment. See J. Maritain (1949) and Glendon (1998). On the compatibility of human rights discourse and "ordinary moral reasoning" (that is, a rejection of the public reason requirement), see Liao and Etinson (2012: 334–336).

removed from the pool of reasons available for public justification. According to some, this removal makes political liberalism the more inclusive philosophy. It does so because, by justifying political action with reference to facts about which reasonable people disagree, the state commits itself to the falsity of certain views, and this commitment excludes the citizens who hold those views.

I have argued that this political liberal position is mistaken about the role of facts in the respectful state's attitude towards its citizens. That someone falsely but reasonably believes that not- p does not remove the fact that p from the pool of available reasons. Rather, it constitutes an additional fact about her to which the state must respond appropriately if it is to respect her.⁴⁶

References

- Arendt, H. (1977). "Truth in Politics," *Between Past and Future: Eight Exercises in Political Thought*. New York: Penguin, 227–264.
- Brandt, R. (1979). *A Theory of the Good and the Right*. Oxford: Oxford University Press.
- Cohen, J. (2008). "Truth and Public Reason," *Philosophy & Public Affairs* 37: 1–42.
- Darwall, S. (1977). "Two Kinds of Respect," *Ethics* 88: 36–49.
- Darwall, S. (2004). "Respect and the Second-Person Standpoint," *Proceedings and Addresses of the American Philosophical Association* 78: 43–59.
- Darwall, S. (2009). *The Second-Person Standpoint*. Cambridge, MA: Harvard University Press.
- Dworkin, R. (2000). *Sovereign Virtue*. Cambridge, MA: Harvard University Press.
- Enoch, D. (2013). "The Disorder of Public Reason," *Ethics* 124: 141–176.
- Estlund, D. (2008). *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.

⁴⁶ I am grateful for comments to Zosia Stemplowska, John Gardner, the participants in the seventh Oxford Studies in Political Philosophy workshop at Syracuse, and an anonymous reviewer for this publication. In particular, I thank my discussant at Syracuse, Sameer Bajaj, for his extremely detailed and perceptive commentary.

- Estlund, D. (2012). "The Truth in Political Liberalism," J. Elkins and A. Norris (eds.) *Truth and Democracy*. Philadelphia, PA: University of Pennsylvania Press, 251–271.
- Frankfurt, H. (1997). "Equality and Respect," *Social Research* 64: 3–15.
- Gaus, G. (2011). *The Order of Public Reason*. Cambridge: Cambridge University Press.
- Glendon, M. A. (1998). "Knowing the Universal Declaration of Human Rights," *Notre Dame Law Review* 73: 1153–1190.
- Hodgson, L.-P. (2010). "Kant on the Right to Freedom: A Defence," *Ethics* 120: 791–819.
- Larmore, C. (1999). "The Moral Basis of Political Liberalism," *Journal of Philosophy* 96: 599–625.
- Leland, R. J. (2019). "Civic Friendship, Public Reason," *Philosophy & Public Affairs* 47: 72–103.
- Liao, S. M. and A. Etinson (2012). "Political and Naturalistic Conceptions of Human Rights: A False Polemic?" *Journal of Moral Philosophy* 9: 334–336.
- Maritain, J. (1949). "Introduction," UNESCO (ed.) *Human Rights: Comments and Interpretations*. New York: Columbia University Press.
- Maritain, J. (1961). "Truth and Human Fellowship," *On the Use of Philosophy: Three Essays*. Princeton, NJ: Princeton University Press.
- Mill, J. S. (1989). *On Liberty and Other Writings*, ed. S. Collini. Cambridge: Cambridge University Press.
- Nussbaum, M. (2003). "Political Liberalism and Respect: A Reply to Linda Barclay," *SATS—Nordic Journal of Philosophy* 4: 24–44.
- Nussbaum, M. (2011). "Perfectionist Liberalism and Political Liberalism," *Philosophy and Public Affairs* 39: 3–45.
- Pallikkathayil, J. (2010). "Deriving Morality from Politics: Rethinking the Formula of Humanity," *Ethics* 121: 116–147.
- Pogge, T. (1997). "Is Kant's *Rechtslehre* Comprehensive?" *Southern Journal of Philosophy* 36: 161–187.
- Quong, J. (2010). *Liberalism Without Perfection*. Oxford: Oxford University Press.
- Quong, J. (2016). "Disagreement, Equality, and the Exclusion of Ideals: A Comment on *The Morality of Freedom*," *Jerusalem Review of Legal Studies* 14: 135–146.

- Quong, J. (2018). "Public Reason," *Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), ed. Edward N. Zalta, URL = <<https://plato.stanford.edu/archives/spr2018/entries/public-reason>>.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rawls, J. (1993). *Political Liberalism*. New York: Columbia University Press.
- Raz, J. (1994a). "Facing Diversity: The Case of Epistemic Abstinence," *Ethics in the Public Domain*. Oxford: Oxford University Press, 60–96.
- Raz, J. (1994b). "Liberalism, Scepticism, and Democracy," *Ethics in the Public Domain*. Oxford: Oxford University Press, 97–124.
- Raz, J. (1998). "Disagreement in Politics," *American Journal of Jurisprudence* 43: 25–52.
- Raz, J. (2016). "Comments on *The Morality of Freedom*," *Jerusalem Review of Legal Studies* 14: 169–181.
- Sinclair, T. (2018). "The Power of Public Positions: Official Roles in Kantian Legitimacy," D. Sobel, P. Vallentyne, and S. Wall (eds.) *Oxford Studies in Political Philosophy Vol. 4*. Oxford: Oxford University Press, 28–52.
- Stemplowska, Z. and A. Swift (2018). "Dethroning Democratic Legitimacy," D. Sobel, P. Vallentyne, and S. Wall (eds.) *Oxford Studies in Political Philosophy Vol. 4*. Oxford: Oxford University Press, 3–27.
- Tadros, V. (2016). *Wrongs and Crimes*. Oxford: Oxford University Press.
- Waldron, J. (1999). *Law and Disagreement*. Oxford: Oxford University Press.
- Wall, S. (2014). "Perfectionism, Reasonableness, and Respect," *Political Theory* 42: 468–489.

5

Perfectionist Duties

Collis Tahzib

It is a central tenet of much contemporary liberal political philosophy that the state should be, in some sense, neutral among the competing conceptions of the good life that exist within society.¹ A neutral state, it is often said, is one that does not impose or even promote any particular ethical, religious, or philosophical doctrine and instead simply upholds a fair framework of laws and institutions within which each individual can freely pursue her own conception of the good life. By contrast, perfectionists hold that one function of the state is to encourage citizens to pursue virtuous, meaningful, or worthwhile ways of life and to discourage them from pursuing degrading or hollow ways of life. To do this, perfectionist states will naturally need to take a stand on a variety of ethical questions, in particular the question of what constitutes human flourishing and excellence.

Perfectionists, however, have not generally been clear about whether they take perfectionist interventions to be permissible or mandatory. Is the perfectionist thesis, that is, that the state *may* promote the good life or, more strongly, that it *must* do so? From what they say, most perfectionists seem quite content to defend the weaker permissibility claim. In this chapter, I argue that perfectionists should set their sights higher. I thus defend what I call *duty-based* perfectionism, according to which the promotion of art, culture, virtue, and other aspects of the good life is a *duty* of the state. In Section 1, I introduce the distinction between duty-based and non-duty-based forms of perfectionism, arguing that despite its simplicity this distinction has been neglected and that the majority of perfectionists would appear to endorse non-duty-based perfectionism. In

¹ See, e.g., Larmore (1987), Rawls (2005), Quong (2011), Patten (2012), and Schouten (2019).

Section 2, I make a case for duty-based perfectionism by defending the idea of a right to a flourishing life. In Section 3, I show how going duty-based can help perfectionists rebut a number of powerful objections that have recently been levelled against perfectionism. Finally, in Section 4, I defend duty-based perfectionism against the charge that it violates the public justification principle.

1. Duty-Based and Non-Duty-Based Perfectionism

Perfectionists hold that that one function of the state is to promote human flourishing and the good life. But they have not always been clear about what *kind* of function this is. According to what I call *non-duty-based perfectionism*, it is permissible but not mandatory for states to promote the good life, whereas according to what I call *duty-based perfectionism* it is mandatory for states to promote the good life.

These two forms of perfectionism admit of internal variation. There are various ways of being a duty-based perfectionist, for instance. One could hold (as I shall argue in Section 2.3) that individuals have a right to the conditions for a flourishing life and that this right generates correlative duties on the part of states to promote the flourishing of the citizenry. Or one might hold that perfectionist duties derive from other duties such as the state's duty to promote justice or self-respect.² For instance, perhaps having a warranted sense of self-respect requires a commitment to pursuing a sound conception of the good life, and so the state's perfectionist duties to promote sound ideals of the good life derive from its general duties to promote the self-respect of citizens. Nor is duty-based perfectionism as a political theory wedded to deontology as a moral theory. For a consequentialist might take human flourishing to be part of her theory of the good and thus support duty-based perfectionism—the view that states

² As this point illustrates, the distinction between duty-based and non-duty-based perfectionism is related to yet distinct from the distinction, drawn by Quong (2011: 29–30, 85, 120–6), between “perfectionist justice” and “non-justice-based perfectionism”. These distinctions diverge because not all duties are duties of justice. For instance, a duty-based perfectionist might hold that states have a moral duty to promote impersonal goods (e.g. through preservation of sites of natural beauty or historical significance), but that this is not a duty of justice. One could therefore endorse duty-based perfectionism without endorsing perfectionist justice.

ought to promote human flourishing—on the grounds that states ought to do what promotes the most good.

Similarly, there are a variety of potentially distinct ways of being a non-duty-based perfectionist: one could hold that perfectionist policies are *permissible*, or that they are *legitimate*, or that they are *justified*, or that there is a *presumption* in favour of them, or that there are *reasons* to pursue them—where each of these notions on its own falls short of establishing a *duty* of perfectionism.

Though most perfectionists have not explicitly distinguished between the claim that the state may promote the good and the claim that the state must promote the good, they have tended to state their views in non-duty-based terms. In his book-length defence of perfectionism, George Sher advances the view that “a government *may* legitimately promote the good” (1997: 1), and in a recent paper he says, more explicitly, “although I don’t think any government is under an *obligation of justice* to promote excellence among its citizens, I fully agree . . . that governments often have good *reason* to do just that” (2018: 60–1). In an influential article, Joseph Chan states: “political perfectionism, which is the focus of discussion in this article, says that it is *permissible* for the state to design its political arrangement or policies with the aim of promoting what the state (or those citizens acting on behalf of the state) thinks are worthwhile goods and ways of life” (2000: 35). In his concluding sentence he describes state perfectionism as “desirable, unavoidable and legitimate” (Chan 2000: 42). In *Liberal Perfectionism* Alexandra Couto argues for the view that “the state is *permitted* and *justified* to protect and promote the opportunities for citizens to engage with intrinsically valuable goods on the basis of their intrinsic value” (2014: 33). And Thaddeus Metz defends what he calls “open perfectionism”, which “recognizes that there are conceptions of the good that could be significantly furthered with the use of state resources and that the potential for a majority to live a certain way can, subject to certain constraints, *justify* such use” (2001: 442).

Of course, claiming that perfectionist policies are permissible, desirable, or justified is strictly speaking compatible with the duty-based perfectionist’s more ambitious further claim that such policies are morally required. But the point is that the kinds of claims made in the

previous paragraph fall short of establishing perfectionist duties and are not supplemented and strengthened by further arguments that would establish any such duties, and thus these authors are most plausibly read as endorsing only non-duty-based perfectionism.

One reason for this lowered sense of ambition is, I suspect, dialectical. That is to say, it is neutralists, and in particular political liberals, who have set the tone of the neutrality-perfectionism debate and who have defined that debate as one about whether it is permissible for the state to act in pursuit of perfectionist ideals. Jonathan Quong, for instance, follows John Rawls in formulating the neutrality-perfectionism debate as a debate about “what liberal states can permissibly or legitimately do”, and in particular about whether “it is permissible for a liberal state to promote or discourage some activities, ideals, or ways of life on grounds relating to their inherent or intrinsic value, or on the basis of other metaphysical claims” (Quong 2011: 15–16). This formulation of the debate has, I suspect, put perfectionists on the back foot, and so most perfectionists have been content to show merely that perfectionist policies are legitimate—that they are not *de*-legitimizing—without going the extra step of arguing that perfectionist policies are in fact a necessary condition of state legitimacy or otherwise morally required.

In my view, perfectionists should set their sights higher. They should not be content to show merely that perfectionism is legitimate, or to show merely that there is no general exclusionary neutrality principle that forbids appeal to considerations about the good life and human flourishing when justifying political arrangements. Rather, they should, I think, defend the more ambitious duty-based form of perfectionism, according to which states are not only permitted but also morally required to promote the good life. As Steven Wall—one of the few perfectionists who argues for this stronger duty-based view—puts it, “political authorities *should* take an active role in creating and maintaining social conditions that best enable their subjects to lead valuable and worthwhile lives” (1998: 8).³

³ Another perfectionist who defends perfectionist duties is Raz: see Raz (1986: esp. 400–29).

2. The Case for Duty-Based Perfectionism

In this section, I make a case for duty-based perfectionism. In Sections 2.1 and 2.2, I consider and reject two recent arguments from Wall, who seeks to establish that perfectionist duties derive from the state's general duties to promote self-respect and fairness. I then argue, in Section 2.3, that duty-based perfectionism is grounded in the moral right of citizens to the fair conditions for a flourishing life.

2.1. From Self-Respect to Duty-Based Perfectionism

In a recent paper, Wall briefly argues that the state's perfectionist duties fall out of its general duties to secure and promote the conditions for the warranted self-respect of its citizens (2013a: 460–5). The most interesting and novel premise of Wall's argument is this: “to have a sense of self-respect, and to merit it, a person must be committed to pursuing a sound conception of the good and must care about his character” (2013a: 461). Thus any state that has a duty to secure the conditions necessary for citizens' warranted self-respect will need, in a perfectionist vein, to ensure that citizens are committed to pursuing a sound conception of the good. If this tight conceptual connection between warranted self-respect and pursuit of a sound conception of the good does indeed exist, then this would generate a dialectically effective argument for duty-based perfectionism because even Rawlsian neutralists generally accept that states have a duty to secure and promote the conditions necessary for citizens' possession of a warranted sense of self-respect.⁴

But when Wall seeks to motivate the idea that warranted self-respect requires commitment to a sound conception of the good, he gives,

⁴ Of course, for those who are not already devotees of Rawlsianism, the idea that states have a duty to promote the necessary conditions for citizens to harbour a warranted sense of self-respect may seem far more contrived, and decidedly less intuitive, than the idea that states should help citizens to lead good, flourishing lives. So, unless more is said about the source of the duty to promote the conditions for warranted self-respect, the argument from self-respect to duty-based perfectionism risks justifying the more plausible by reference to the less plausible. (I thank Jeff McMahan for this point.) But here I grant the existence of a duty to promote the conditions for warranted self-respect.

I believe, the wrong kind of case. Wall invites us to consider an individual, Charlie, who “critically evaluate[s] his conception of the good” and “comes to see that it is misguided” (2013a: 463). As Wall plausibly says, Charlie must now renounce his commitment to his conception of the good. Charlie, that is, could not simultaneously enjoy warranted self-respect and judge as unsound the conception of the good life to which he is committed: “in being indifferent to whether or not he, in fact, lives well, Charlie would thereby show that he did not really respect himself” (2013a: 463). Any self-respect in this case would, says Wall, be “deluded”, not “warranted” (2013a: 463). Thus, warranted self-respect “rests on the judgement or conviction that one’s plan of life or conception of the good is worth carrying out” (Wall 2013a: 462).

The problem with the Charlie case is that it does not support the claim that warranted self-respect requires commitment to a conception of the good life that is sound. What it supports is the claim that warranted self-respect requires commitment to a conception of the good life that one *believes* is sound. So, assuming that the state has a duty to promote warranted self-respect, the Charlie case could only support a kind of subjective perfectionism (the state has a duty to promote conditions in which citizens will be committed to conceptions of the good that they believe to be sound), not the objective perfectionism (the state has a duty to promote conditions in which citizens will be committed to conceptions of the good that are *genuinely* sound) that Wall is after.

To test whether warranted self-respect requires commitment to a sound conception of the good, we thus need to consider an individual who justifiably believes that his conception of the good is sound but who is simply mistaken about this. So consider Dan, an activist whose life is structured around campaigning for the reintroduction of the death penalty in England. He has been convinced, we can suppose, by Matthew Kramer’s “purgative rationale” according to which certain crimes are so extravagantly and defilingly evil that the continuation of the evildoer’s life would be a blot on the moral character of the community to which he belongs and would constitute an affront to humanity, and thus the death penalty is called for to purge the community of that taint (Kramer 2011). Dan is admired amongst advocates of the death penalty, and he is often invited to give talks about the ethics of capital

punishment. But suppose also, ex hypothesi, that the death penalty is always impermissible, and thus that Dan's pro-death-penalty stance is mistaken. So Dan is committed to an unsound conception of the good (despite being committed to a conception of the good that he *judges* to be sound).

The question, then, is this: Must Dan—who is committed to a reasonable but ultimately unsound conception of the good—lack warranted self-respect? If he does respect himself, must he be “deluded” in doing so? It is difficult to see why he must. Dan certainly seems to fit the profile of a warrantably self-respecting individual: he cares about his beliefs and thinks they matter enough to be worth subjecting to critical scrutiny, he confidently protests behaviour that he reasonably and justifiably (if mistakenly) takes to be wrongful, he feels a sense of pride and accomplishment when his activism bears fruit, and so on. Of course, we might think that Dan's sense of self-respect would, in some way, be *more* warranted if it were premised on a sound conception of the good. But that is no reason for thinking that his sense of self-respect is unwarranted or deluded. In general, then, it seems that the objective soundness of one's conception of the good is an inordinately strict requirement for having a warranted sense of self-respect.⁵

Let me briefly defend this argument against two objections. First, is it not possible to specify Dan's conception of the good as *being an effective campaigner for political change* rather than as *campaigning to get the death penalty reintroduced*? In this case, the unsoundness of his views about the death penalty would not suffice to render unsound his (more abstractly stated) conception of the good, and so this would no longer be

⁵ It is worth adding here that, even if Dan does not qualify as having warranted self-respect, it is unclear how steering him, in Wall's preferred perfectionist way, into having sound views about capital punishment could help him to enjoy warranted self-respect. Would Dan really have enhanced grounds for self-respect if the only reason he had the right view about capital punishment was that he had been “nudged” by the government? There is thus a gap between establishing that warranted self-respect requires commitment to a sound conception of the good and establishing that states (who have a duty to secure warranted self-respect) should promote sound conceptions of the good, because the state's efforts to promote the good (or at least certain kinds of efforts by the state to promote the good) might themselves compromise the warrants for self-respect. But, as with the concession in footnote 4, I will not press this point further.

a counterexample in which an individual has warranted self-respect despite having an unsound conception of the good.⁶

Tinkering with the level of abstraction at which we state conceptions of the good does not rescue Wall's argument, however. It merely moves the bump in the rug. Thus suppose, *ex hypothesi*, that Buddhism and other apolitical religions are right: we should reject the power struggles, the passions, and the divisive activism of earthly politics in favour of a life of spiritual contemplation. Would the objective unsoundness of the life of politics mean that Dan (who, recall, confidently protests behaviour that he sees as wrongful, feels pride in his accomplishments, and so on) must be deluded to respect himself? Intuitively, it is again unclear why he must.

Second, Wall might argue that the case of Dan only serves to demonstrate that one can be self-respecting whilst having an unsound conception of the good; it does not demonstrate that one can be *warrantedly* self-respecting whilst having an unsound conception of the good. Insofar as having an objectively sound conception of the good is *defined* into the notion of warrantedness, cases such as Dan's, which involve warrantably self-respecting individuals who subscribe to unsound conceptions of the good, are ruled out conceptually.

However, if warrantedness rather than self-respect carries too much of the argumentative load here, the move from warranted self-respect to duty-based perfectionism quickly becomes a tautology—a matter of conceptual fiat rather than an intuitively persuasive argument. Of course, warrantedness can play *some* role in Wall's argument: it is plausible that someone who is systematically deluded (e.g. someone in Nozick's "experience machine" who takes pride in illusory accomplishments), or who is guided by a deeply objectionable ideology (e.g. a racist who takes great pride and satisfaction in his racist behaviour), is not warrantably self-respecting, however many of the hallmarks of self-respect he might possess. But if warrantedness does too much moral work beyond this, much of the intuitive force of the value of warranted self-respect is lost. In such a case, those who are not already convinced by the idea of perfectionist duties are likely to respond that, while they care about

⁶ I thank David Miller for putting this objection to me.

warranted self-respect in *some* sense (in particular, where warrantedness plays only a fairly minimal role), they simply do not care about *this* specific account of warranted self-respect—an account that feels very custom-built, even ad hoc, and that functions like a kind of Trojan horse.

2.2. From Fairness to Duty-Based Perfectionism

In another recent discussion, Wall has advanced a brief yet interesting argument that the state's duties to promote the good life derive from its general duties of fairness (2013b: 98–100).⁷ Wall first argues that “fairness sometimes requires that, holding other things constant, we should favor those who are less well off over those who are better off” (2013b: 99). Wall then argues that “moral environments that offer their inhabitants bad options, when compared to available alternatives, will tend to disadvantage those who are less well off” (2013b: 99). As Wall also puts this second premise: “the failure to engage in [perfectionist politics] comes at the expense of those who fare least well” (2013b: 99). So since fairness requires priority for the worst off, and since the worst off are particularly disadvantaged by liberal neutrality, fairness requires perfectionist policies and interventions.

Libertarians will want to reject Wall's first premise—namely, that fairness requires priority for the worst off. But my worry is with his second premise: that the worst off are particularly disadvantaged by neutrality (and so are particularly advantaged by perfectionism). To be clear: I believe that neutrality does disadvantage citizens by cutting them off from important goods. What I am questioning is Wall's claim that neutrality is *differentially* disadvantageous to the worst off, such that it would raise concerns about fairness.

So what reason does Wall give for thinking that a society that does not enact perfectionist policies would be particularly disadvantageous to the worst off? His main reason seems to be that those who are worse off are worse decision-makers and thus are more vulnerable to bad options. He

⁷ Wall's argument draws on a similar argument from fairness to paternalism developed in Arneson (1989).

says: “Naturally, and speaking generally, those who are bad decision-makers will tend to be less well off overall than those who are good decision-makers” (Wall 2013b: 99). Thus a society that conducts various state-run lotteries, for instance, “will disadvantage those who are not good at making judgements of probability, or have trouble resisting the impulse to gamble”, as compared to those who are able to see through such schemes (Wall 2013b: 99). A perfectionist politics, which removes this option, will thus benefit the worse off (who are worse decision-makers) more than the better off (who are better decision-makers). Wall also gives the example of heroin abuse. The existence of this option disadvantages the worse off and so the family of a heroin addict might reasonably question how fair it is that “the moral environment in which he lived made it easier than it need be for people like him to ruin their lives” (Wall 2013b: 100).

Wall may well be correct that the worse off are disproportionately disadvantaged by the existence of worthless options such as drugs and lotteries. But, in general, perfectionism does not merely involve the *removal of worthless options*; a major part of the perfectionist programme is the *provision of worthwhile options*. By Wall’s own logic, it would seem that the better off (insofar as they are better decision-makers) will be more likely to take advantage of these options, whereas the worse off (insofar as they are worse decision-makers) will be more likely to ignore them. Consider, for instance, Quong’s list of standard perfectionist policies: “performance art, art galleries, public parks, works of literature, sights of cultural significance, education programmes for adults, and athletic events” (2011: 89). On the face of it, these kinds of options seem to cater to middle-class pastimes and are likely to be most advantageous to the better off.

So once a broad range of perfectionist policies is on the table—both policies that remove worthless options and policies that provide worthwhile options—it is not clear why we should think that any particular segment of society would stand to gain more or less from perfectionism than any other. Some perfectionist policies will likely be of most benefit to the worse off; others will likely be of most benefit to the better off. How do all these benefits wash out? It is difficult, if not impossible, to say from the armchair. What is needed here is some robust, empirically informed sociological model that explains why we should expect the worse off to

be greater beneficiaries of perfectionist policies. In the absence of this, the argument from fairness to duty-based perfectionism does not go through.

2.3. A Right to a Flourishing Life

In my view, perfectionist duties follow not from duties of self-respect or fairness but rather from the moral right of all citizens to lead a flourishing life. This moral right correlates with the government's perfectionist duty to promote the good life.

A clarification is immediately in order: although the notion of a "right to a flourishing life" is a useful shorthand, what is really being guaranteed by this right is not a flourishing life *itself*. One important reason for this is that a right to lead a flourishing life would not be responsibility-sensitive. Such a view would be incompatible with the idea that individuals should take personal responsibility for their conception of the good life and should bear the costs and burdens associated with its pursuit. Put differently, a right to a flourishing life itself would end up indulging some citizens' expensive tastes. Instead, then, what the right to a flourishing life guarantees is the *fair conditions* for living a flourishing life, understood to be the fair set of opportunities, resources, public culture, and other conditions promotive of and conducive to the flourishing of citizens generally. This view is responsibility-sensitive because what citizens choose to do once these fair conditions have been provided is up to them.

But what justifies the claim that there does in fact exist a right to the fair conditions of human flourishing—or, as I will put it elliptically, a right to a flourishing life? While many are willing to accept the idea of a right to a "minimally decent life", it is not generally thought that there is a right to a flourishing life.

To justify this claim, I want to consider whether an original position heuristic might yield a right to a flourishing life.⁸ Here we determine the scheme of rights and duties by considering what social terms would be

⁸ I explore the set-up of the "perfectionist original position", and the reasoning that leads parties in this original position to perfectionist principles of justice, in much greater depth in my book-in-progress entitled *A Perfectionist Theory of Justice*.

chosen by rational and reasonable contractors situated behind a veil of ignorance. In particular, we want to know whether parties in the original position would select social terms that are deeply shaped by perfectionist considerations and that are promotive of and conducive to moral, intellectual, and artistic excellence.

In *A Theory of Justice*, Rawls discusses some proposals along these lines (1971: 325–32). He first considers an extreme Nietzschean form of perfectionism according to which rights and duties are defined in such a way as to “maximize the achievement of human excellence in art, science, and culture” (Rawls 1971: 325). As Rawls rightly notes, this extreme and maximizing form of perfectionism would be rejected by parties in the original position because (inter alia) it “might lead to a lesser religious or other liberty, if not to a loss of freedom altogether to advance many of one’s spiritual ends” (Rawls 1971: 327).

But this still leaves open a second and more moderate form of perfectionism. Just as Rawls imagines a “mixed conception” of utilitarianism in which the principle of utility is subordinated to the equal liberties principle and the fair equality of opportunity principle, we can imagine what I shall call a “mixed conception of perfectionism” in which a principle of perfection is subordinated to other lexically prior principles of justice. This principle of perfection would require the establishment and maintenance of social conditions promotive of and conducive to flourishing ways of life. On this mixed conception, once everyone has equal basic liberties and once everyone has fair equality of opportunity, social resources are intuitionistically balanced between further pursuit of equality (as dictated by the difference principle) and pursuit of moral, intellectual, and artistic excellence (as dictated by the principle of perfection). There can be stronger and weaker forms of this mixed conception of perfectionism, depending on how much weight is assigned to the principle of perfection as compared to the difference principle during the intuitionistic balancing stage. But even the strongest form of the mixed conception of perfectionism would respect basic rights and liberties as well as fair equality of opportunity.

The parties in the “perfectionist original position”, as we can call it, would therefore be presented with a menu that includes the following perfectionist principles of justice:

- (1) Each person has the same indefeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all [the equal basic liberties principle];
- (2) Social and economic inequalities are to be attached to offices and positions open to all under conditions of fair equality of opportunity [the fair equality of opportunity principle];
- (3) Social and economic inequalities are to be to the greatest benefit of the least-advantaged members of society [the difference principle]; and
- (4) Social conditions promotive of and conducive to flourishing ways of life are to be established and maintained [the principle of perfection],

where (1) is lexically prior to (2), (2) is lexically prior to (3) and (4), and (3) and (4) are not ordered in terms of lexical priority and must instead be intuitionistically balanced against each other.⁹

It seems plausible to suppose that contractors would select this mixed conception of perfectionism in which the claims of flourishing and excellence are weighted with moderation. In addition, that is, to selecting social terms that guarantee equal basic rights and liberties, that require positions and offices to be allocated on the basis of fair equality of opportunity, and that give special distributive priority to the least-advantaged members of society, it is difficult to see why parties in the original position would not also think it reasonable and prudent to select social conditions that are promotive of and conducive to their moral, intellectual, and artistic excellence.

Rawls explains that his primary reason for forgoing “any further advantages that might be won” by going beyond the two principles of justice is that “the hardship if things turn out badly [may be] intolerable” (1971: 175). From the point of view of the original position, trying to win any further advantages beyond Rawls’s principles is too risky for contractors. But what are the risks involved in selecting the mixed conception of perfectionism? Given that the principle of perfection is conditional on lexically prior principles of justice having been met,

⁹ For principles (1)–(3), see Rawls (2001: 42–3).

there is no risk that this view will “lead to a lesser religious or other liberty, if not to a loss of freedom altogether to advance many of one’s spiritual ends” (Rawls 1971: 327). A gamble in which some citizens might undermine their religious freedoms is simply not on the cards for proponents of the mixed conception of perfectionism. Nor is there a risk that this view will lead to great inequalities: the equal liberties principle and the fair equality of opportunity principle are quite demanding principles that would significantly constrain permissible social and economic inequalities, and the influence of the difference principle at the intuitionistic balancing stage would constrain these inequalities even further. In short, the stakes are much lower under the mixed conception of perfectionism because the lexically prior principles set a kind of social minimum; it is difficult to see who could reasonably find it an intolerable hardship to live under a system that, in addition to upholding Rawls’s two principles of justice, seeks to foster social conditions that encourage citizens to lead flourishing lives.

Rawls is aware of this more moderate kind of view and admits that it is “not easy to argue against” (1971: 330). In an interesting anticipation of his later work, Rawls’s main worry is about the disagreement and indeterminacy surrounding perfectionist considerations:

Criteria of excellence are imprecise as political principles, and their application to public questions is bound to be unsettled and idiosyncratic, however reasonably they may be invoked and accepted within narrower traditions and communities of thought . . . Appeal to perfectionist criteria [is often made] in an ad hoc manner. When it is said, for example, that certain kinds of sexual relationships are degrading and shameful, and should be prohibited on this basis, if only for the sake of the individuals in question irrespective of their wishes, it is often because a reasonable case cannot be made in terms of the principles of justice. Instead we fall back on notions of excellence. But in these matters we are likely to be influenced by subtle aesthetic preferences and personal feelings of propriety; and individual, class, and group differences are often sharp and irreconcilable. Since these uncertainties plague perfectionist criteria and jeopardize individual liberty, it seems best to rely entirely on the principles of justice which

have a more definite structure. Thus even in its intuitionistic form, perfectionism would be rejected as not defining a feasible basis of social justice. (Rawls 1971: 330–1)

I address in Section 4 the charge that perfectionist considerations are too controversial or indeterminate to play a role in politics, so my full response to this worry must await that section. Nonetheless, let me make three specific comments here on this important passage from Rawls.

The first point is that although Rawls is meant to be considering a more moderate and intuitionistic kind of perfectionism at this point, he talks as if the perfectionism is far more extreme and Nietzschean in kind. For instance, he suggests that the perfectionist view in question may “jeopardize individual liberty” and that it may say that certain kinds of relationships “should be prohibited” on the grounds that they are degrading. But, to repeat, on the mixed conception of perfectionism that I am proposing, the provision of an environment that is promotive of and conducive to moral, intellectual, and artistic excellence is *conditional* on the other lexically prior principles of justice being met (including, in particular, the principle that “each person has the same infeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all”). So prohibited relationships and jeopardized liberty are not in fact risks of the mixed conception of perfectionism rightly understood.

The second point is that Rawls’s consideration of one specific class of perfectionist judgements (namely, judgements about sexual morality) is insufficient to establish some *general asymmetry* between the controversy or indeterminacy of perfectionist and non-perfectionist judgements, such that the latter but not the former could shape social terms.¹⁰

On the one hand, many perfectionist judgements are *not* deeply controversial or indeterminate. For instance, it seems likely that very many people—at least as many people as would agree to any comparably specific non-perfectionist judgement—would agree that the life of

¹⁰ This point is connected to the so-called asymmetry objection to political liberalism. See, e.g., Caney (1998), Sandel (1998: 202–10), and Chan (2000).

someone who is wholly consumed by drug addiction is less meaningful and worthwhile than the life of someone who reads philosophy and literature, has close friends, and has an upright character. More generally, it is unclear why appeal to a core handful of perfectionist values that have broad appeal and that are pitched at a high level of generality and abstraction (such as moral, intellectual, and artistic excellence) would be too “ad hoc”, “idiosyncratic”, or “uncertain” to feasibly be part of social justice.¹¹ So while Rawls may well be correct that judgements about sexual morality are too idiosyncratic or controversial to be a feasible basis of social justice, it is doubtful that this is true of perfectionist judgements generally.

On the other hand, many non-perfectionist judgements *are* deeply controversial or indeterminate. For instance, many people disagree heatedly about what fair equality of opportunity requires in relation, say, to employment practices. For some, it requires only non-discrimination along morally arbitrary lines; for others, it requires us to “wrestle with counterfactuals about what the [applicants] would have been like if they had had each other’s backgrounds” (Richards 1998: 56); and for yet others, it requires us to take into account the unearned effects not only of the “social lottery” but also of the “genetic lottery”. So deep are these disagreements that some wonder whether there is any remaining shared ground: “the fact that we all converge on a certain form of words [i.e. ‘fair equality of opportunity’] does not mean we actually agree on anything substantial” (Cavanagh 2002: 1–2).

In short, Rawls’s brief consideration of a single example does not suffice to establish that the distinction between perfectionist and non-perfectionist judgements tracks the distinction between what is and what is not a matter of intractable controversy and indeterminacy.

The third point is that, as Rawls himself says, it is impossible to make highly precise judgements and measurements of non-perfectionist values such as freedom, fairness, equality, and self-respect. Our judgements

¹¹ Of course, these abstract perfectionist values will become more controversial as they are made more concrete. But the same is true of abstract liberal values such as equality and freedom. So to insist that the concretization of abstract perfectionist values will incur more controversy than the concretization of abstract liberal values does not *vindicate* the asymmetry claim and instead only *relocates* that claim.

“[need] not be very exact”, Rawls explains, and only need to be “accurate enough to guide the main decisions” concerning the provision of fair social terms (1971: 327). So the question is: Can we make *sufficiently* determinate assessments about the good life to guide certain important political decisions? Yet, as Rawls earlier concedes, many (though of course not all) perfectionist judgements appear to pass this test: “clearly there are standards in the arts and sciences for appraising creative efforts” and so “comparisons of intrinsic value can obviously be made” (Rawls 1971: 328). “They are not necessarily so vague,” Rawls continues, “that they must fail as a workable basis for assigning rights” (1971: 328).

Pulling together these three points, Rawls’s reply gives us no reason to doubt the central claim of the mixed conception of perfectionism. Rawls’s reply, that is, gives us no reason to doubt that *certain* abstract and widely shared perfectionist values such as moral, intellectual, and artistic excellence (as opposed to highly specific claims about sexual morality) are *sufficiently* precise and determinate (as opposed to perfectly precise and determinate) to shape social terms in a manner *consistent* with the other principles of justice (as opposed to doing so in a manner that jeopardizes liberties and prohibits relationships).

2.4. Indistinguishability?

At this point, however, duty-based perfectionism might be thought to face a serious objection. For liberal neutralists can ask:

In what sense does the neutral liberal state not *already* secure the fair conditions for human flourishing? After all, a Rawlsian well-ordered liberal society would eradicate the kind of poverty that prevents many individuals from devoting their time and energies towards the best things in life; it would combat non-material obstacles to flourishing such as disinformation, propaganda and discriminatory speech; it would encourage the development and exercise of the two moral powers, one of which is a capacity to rationally form, revise and pursue a conception of the good; it would provide space for individuals to interact and form associations aimed at various forms of excellence.

What is involved in providing the fair conditions for human flourishing above and beyond this? Once each citizen has been given her fair share of social and economic goods, why hasn't she also been given the fair conditions necessary for her to flourish? What does duty-based perfectionism do that would not already be done by liberal neutrality?¹²

Indeed, Quong rejects duty-based perfectionism and focuses almost entirely on non-duty-based perfectionism precisely on the grounds that duty-based perfectionism cannot "practically distinguish itself from non-perfectionist theories of distributive justice, such as Rawls' or Ronald Dworkin's theory" (2011: 122).

For duty-based perfectionists, however, the prescriptions of Rawls, Dworkin, and other liberal neutralists fall short of providing fair conditions for flourishing because perfectionist considerations strongly influence one's conception of social or distributive justice. The notions of "fair shares", "fair terms", and "fair conditions" are deeply (indeed, constitutively) shaped by specific claims about the good life and human flourishing. Perfectionist considerations, in other words, *define* what constitutes fair shares and fair conditions. So whereas many perfectionists (of the non-duty-based kind) speak as though perfectionism is the cherry on top—the policies added on, once a state has established justice and the fair terms of social cooperation—duty-based perfectionists embed considerations of human flourishing more deeply into their account of political morality. To ask what more could be required over and above the establishment of the Rawlsian fair terms of social cooperation is thus to miss the point because, for the duty-based perfectionist, what counts as fair terms and what counts as the rightful treatment of the citizenry by the state is precisely what is at stake here.¹³

¹² I thank Max Afnan and Jonathan Quong for pressing me on this point.

¹³ I thank Jonathan Quong and Steven Wall for helping me to formulate this response. There are two other available responses to the indistinguishability objection. First: even if duty-based perfectionism is extensionally equivalent to the leading non-perfectionist theories of justice, this does not show that duty-based perfectionism is redundant or uninteresting, because they might agree for different reasons. Duty-based perfectionism might, that is, give a more attractive underlying picture of political morality, one that justifies a common set of policies in a more natural and plausible way. Consider Rawls's response (1971: 332) to the worry that his view seems to cast doubt on the moral permissibility of state support for the arts and culture: "The principles of justice do not permit subsidizing universities and institutes, or opera and the

To illustrate: suppose, as I have been saying, that moral, intellectual, and artistic excellence are components of a flourishing life. Then, for duty-based perfectionists of this kind, the natural and plausible thing to say is that social conditions cannot be fair unless they foster and promote moral, intellectual, and artistic excellence. For such duty-based perfectionists, then, providing the fair conditions of flourishing would call for more than the liberal neutralist's efforts to secure associational liberties, to eradicate poverty, and to combat discriminatory speech: it would also require the cultivation and promotion of moral, intellectual, and artistic excellence. These perfectionist ideals could be expected to heavily shape policy in a large range of areas including "public assistance, educational policy, the criminal and civil justice system, the prison system, city planning and land use, transportation policy, the tax code, support for cultural institutions, regulation of the entertainment industry, investment incentives, and the structure of institutions such as the military", to employ Sher's helpful list (1997: 246).

One might still worry that the perfectionist ideals of moral, intellectual, and artistic excellence are pitched at such a high level of generality that it is difficult to picture the concrete policies that these ideals would support in areas such as city planning, transportation, or the military. What does duty-based perfectionism look like in even more concrete terms? What is a specific example of a policy that would be enacted by duty-based perfectionism but that non-perfectionist theories of justice could not support? The worry is that until we have an answer to this question—until we work out what duty-based perfectionism looks like at a lower level of abstraction and until we have in clear view some specific policy or policies that duty-based perfectionism allows but that other

theatre, on the grounds that these institutions are intrinsically valuable... Taxation for these purposes can be justified only as promoting directly or indirectly the social conditions that secure the equal liberties and as advancing in an appropriate way the long-term interests of the least advantaged." This strikes me as an implausibly strained and contrived justification for the arts—a clear case in which non-perfectionist theories support the right policy for the wrong kind of reason. Second: even if duty-based perfectionism is, to a large extent, both extensionally and intensionally indistinguishable from non-perfectionist views, it is still unclear why this should be an objection to perfectionists rather than non-perfectionists, since indistinguishability is a symmetrical relation. Why not think that it is *neutralists* who are aping the good-focused political theory of perfectionists? But since I believe there will be genuine divergences at the bottom line between duty-based perfectionism and non-perfectionist theories of justice, I shall not press these points further.

views cannot sign onto—it is difficult to get a feel for the ways in which duty-based perfectionism comes apart from leading non-perfectionist theories of justice.

I am reluctant to provide granularity of this kind because I am concerned with the general question of whether the state should seek to promote the good life; I am less interested in the more specific question of what particular *conception* of the good life the state should promote, or the even more specific question of what particular *policies* the state should use to promote a particular conception of the good life. Giving highly specific examples is thus hazardous because it invites anti-perfectionists to focus on and quibble with the examples rather than considering the moral bearings of perfectionism as a general matter and in some or another form. Indeed, my suspicion is that the relative unpopularity of perfectionist political philosophy can be accounted for in precisely these terms. On the one hand, this is the doing of proponents of perfectionism, who have not typically theorized perfectionism at the levels of generality that best demonstrate its philosophical appeal and flexibility and who have instead moved hastily to the defence of particular and often fairly idiosyncratic varieties of perfectionism; on the other hand, this is the doing of critics of perfectionism, who frequently speak as if the shortcomings of particular instances of perfectionism (e.g. highly conservative religious versions of perfectionism) impugn perfectionism as a general thesis.

Still, I recognize that examples do help to fix ideas and to make a debate come alive. So here might be one concrete policy difference: a duty-based perfectionism of the kind I have sketched might plausibly seek to regulate the profusion of commercial advertisements that seek to instil in masses of people an endless and ever-changing array of enthusiasms, thereby directing energies away from various forms of excellence and towards the feverish accumulation of material possessions. While anti-perfectionists also support some regulations on advertisements (for instance, the prohibition of false or misleading claims), a duty-based perfectionism based on moral, intellectual, and artistic excellence would likely support more extensive regulations (though of course any such regulations would need to be sensitive to other values, such as freedom of speech). On this duty-based perfectionist view, then, fair conditions

could be expected to go considerably beyond those that would be established within a well-ordered liberal-neutral society.

Without these kinds of conditions—conditions that are shot through with perfectionist values—it is not, of course, *impossible* for citizens to realize moral, intellectual, and artistic excellence. Under liberal neutrality, citizens clearly still have some chance of attaining moral, intellectual, and artistic excellence. Nevertheless, as far as this form of duty-based perfectionism is concerned, without these kinds of conditions citizens are being denied a *fair* chance to lead a flourishing life and so they have a legitimate complaint against the state. In this specific sense, and to this extent, duty-based perfectionism entails that even a well-ordered, perfectly realized Rawlsian state fails to respect some of the moral rights of its citizens.

None of this, yet, is to say that duty-based perfectionism is in the clear. Perhaps, insofar as it appeals to controversial claims about excellence, duty-based perfectionism fails to take seriously the reasonable pluralism within a free society and so contravenes the public justification test—an objection I consider in Section 4. The point is just that the move from non-duty-based to duty-based perfectionism need not involve the abandonment of perfectionism: by building perfectionist content into one's conception of fair terms, as duty-based perfectionists will find it natural to do, it is possible to overcome Quong's charge that a duty-based form of perfectionism aimed at the provision of fair conditions for flourishing would no longer be distinguishable from liberal neutrality.

3. The Advantages of Going Duty-Based

The stronger duty-based form of perfectionism has the advantage of giving perfectionists the resources to block a number of powerful recent objections. I shall illustrate this by examining three recent objections which allege respectively that perfectionism is illegitimate, that it is paternalistic, and that it manifests a meddlesome “quidnunc” mentality.¹⁴ All of

¹⁴ Quong's (2011: 45–72) autonomy-based objection to perfectionism also presupposes that there are no perfectionist duties and as such would be powerless against duty-based

these objections, however, presuppose there are no perfectionist duties—a fair assumption in relation to mainstream non-duty-based forms of perfectionism, but one which flatly begs the question against duty-based perfectionists. Insofar as these objections trade on the absence of duties to promote the good life, then, duty-based perfectionism is better placed to rebut them than are non-duty-based forms of perfectionism which recognize only reasons, justifications, and permissions to promote the good life.

3.1. Legitimacy

In his *Liberalism Without Perfection*, Quong (2011: 108–36) argues that perfectionist states lack legitimacy. To make this argument, he notes that the legitimacy of perfectionist states is generally defended by reference to “practical reason” theories of legitimacy. According to these theories—of which Raz’s “service conception” is the foremost example—states are legitimate when they enable us to best comply with the reasons that apply to us. A surgeon who comes across a serious accident, for instance, has legitimate authority to issue commands to a bystander with no medical expertise because that bystander will best comply with the reasons that apply to her by following the surgeon’s commands rather than by deciding for herself how to assist the victim (Quong 2011:

perfectionism. For Quong, the perfectionist state undermines the autonomy of citizens because it coercively takes their money and spends it on what it judges to be valuable (e.g. art galleries) rather than letting the money remain with citizens to spend as they see fit (e.g. buying a new games console). The perfectionist state thus “subject[s] the will of citizens to its own perfectionist judgement” (Quong 2011: 65). But one could imagine someone running this same line of argument against standard liberal policies such as public education. Doesn’t the state undermine citizens’ autonomy when it taxes them in order to fund public education? Doesn’t the liberal state subject the will of citizens to its own liberal judgement? Quong would likely reply that being forced to pay taxes towards public education is not autonomy-undermining because we are under a *duty* to contribute towards the provision of public education; we never had any right in the first place to the money that the liberal state takes from us and uses to subsidize public education. But if one is a *duty-based* perfectionist, then one will want to say the same thing about perfectionist policies such as art galleries: being forced to pay taxes towards art galleries is not autonomy-undermining because we are under a duty to contribute towards the provision of art galleries; we never had any right in the first place to the money that the perfectionist state takes from us and uses to subsidize art galleries. This point is also made in Kramer (2017: 59–62).

112–13). On practical reason views, then, perfectionist states are legitimate because they enable citizens to flourish and thus to comply with the full range of practical reasons that apply to them.

But, says Quong, these practical reason models of legitimacy all suffer from a common problem: namely, that the mere fact that living by X's rules will help you to live well and to comply with the dictates of reason does not suffice to show that X has legitimate authority over you. We do not, for instance, think that just because a certain Peruvian tour company is superb in every way—so good, in fact, that what you have most reason to do is to become its customer—it thereby gains legitimate authority over you and may force you to sign up to its tours when you visit Peru (Quong 2011: 108–9). The “fundamental problem” with the practical reason accounts of legitimacy that have tended to underwrite perfectionist politics, then, is that they fail to explain “why the brute fact that *I* have reason to do something should affect what rights *you* have with regard to me” (Quong 2011: 115).

In light of this, Quong develops his own preferred “natural-duty” account of legitimacy, according to which legitimate authority depends on *rights and duties, not reasons*. As Quong puts this view, “legitimate authority does not track what we have reason to do, it tracks what rights we have, and what duties we may be under” (2011: 116). This view explains the Peruvian tour company case: the tour company has no authority over me because they have no *right* to my custom and because I have no corresponding *duty* to use their services (even though I do have *reason* to do so). And the natural-duty account also offers an alternative explanation of the surgeon case: the surgeon has legitimate authority over me because I am under a positive duty to aid and can best fulfil this duty by accepting the surgeon's commands as authoritative (and *not*, as Raz would say, because by following the surgeon's commands I best comply with the reasons that apply to me). Of course, this is not to say that talk of rights and duties has no connection to talk of reasons. Indeed, what we have most reason to do and what we have a duty to do will often coincide. But, as the Peruvian tour case illustrates, these can come apart because rights and duties are determined by a set of reasons that is narrower than the full set of practical reasons, which includes friendship, humour, wisdom, and many other values (Quong 2011: 119).

In short, then, legitimate authority depends on the ability of states to “enable each of their citizens to better fulfill the duties they are under”, and not on their ability to help us lead the kind of flourishing lives that we have most reason to lead (Quong 2011: 131). This is why, for Quong, neutral states, whose policies aim at ensuring citizens fulfil their duties of justice to others, are legitimate, whereas perfectionist states, whose policies aim at ensuring citizens live flourishing lives in accordance with the full set of practical reasons that apply to them, are not legitimate.

However, this objection assumes that there are no perfectionist duties. Quong’s argument thus begs the question against duty-based perfectionists. After all, if, as I have argued, each citizen has a right to flourish, and if the state and other citizens therefore have correlative duties to promote such flourishing, then the legitimacy of the perfectionist state could be grounded in these perfectionist duties. In such a case, it could be said not just of neutral states but also of perfectionist states that they possess legitimacy in virtue of planning and coordinating individual actions in ways that “enable each of their citizens to better fulfill the [perfectionist] duties they are under”. Quong’s preferred natural-duty account of legitimacy would, that is, apply equally well to perfectionism as to neutralist liberalism. It is important to note that Quong (2011: 120–6) is aware of this possibility and does consider a perfectionist response along these lines—though he is not ultimately convinced for reasons I discuss in Section 2.4. But, for our purposes here, the crucial point is that the cogency of the legitimacy objection to perfectionism entirely hinges on, and reduces to, the claim that there are no perfectionist duties.

3.2. Paternalism

Quong also contends that perfectionism is objectionably paternalistic. To show this, he first argues for a “judgemental” account of paternalism according to which A acts paternalistically when she attempts to improve the welfare, good, interests, or values of B, and when her act is motivated by a negative judgement about B’s ability to make the decision that would effectively advance B’s welfare, good, interests, or values (Quong 2011: 80). This account of paternalism, says Quong, “captures our sense

that to treat someone paternalistically is to treat that person like a child in the specific sense of acting in that person's best interests because you believe, in this situation, the person lacks the ability to do so himself or herself" (2011: 81). And this account explains the *prima facie* wrongness of paternalism. For paternalism conflicts with the conception of moral status central to liberal political philosophy, according to which citizens are free and equal in virtue of their possession of two moral powers, namely the capacity for a sense of justice and the capacity to form, to revise, and rationally to pursue a conception of one's good. Thus to treat someone paternalistically is *prima facie* wrong because it involves "treat [ing] that person as if he or she lacks the second moral power" (2011: 101).

Quong then seeks to demonstrate the paternalism of perfectionism by arguing that perfectionists have no satisfactory answer to a "simple but important" question: "Why is state action necessary at all to achieve perfectionist objectives?" (2011: 74). "Why not simply give each citizen their fair share of resources and let them make their own decisions [about how to lead a flourishing life]?" (Quong 2012: 2). Quong considers various answers given by perfectionists to this question and finds nearly all to involve a negative, infantilizing judgement about citizens' capacity to run their own lives, thus revealing "perfectionism's true paternalist colours" (2011: 86). For instance, some perfectionists have argued that the realization of ideals of the good life requires political action because, when left to their own devices, people act *irrationally* and fall prey to various cognitive biases that lead them to squander their resources. Other perfectionists, Quong notes, claim that citizens suffer from *weakness of will* and thus public action is required in order to steer them away from unworthy temptations and addictions.

In responding to the charge that there is no non-paternalistic justification for perfectionist public action, it is, I think, helpful to consider how neutralist political liberals might themselves respond to this same objection. Why, that is, is state action necessary at all to achieve *ordinary, non-perfectionist, justice-based* objectives? Is it not similarly paternalistic for political liberals to use state power to pursue neutral ends such as education, national defence, and social justice? Do political liberals have non-paternalistic grounds for believing that these neutral ends would not

be achieved in the absence of state action? In other words: Why think that perfectionist state action is paternalistic and non-perfectionist state action non-paternalistic?

Quong says little about this question, but in a footnote he does suggest that state action in areas such as social justice and national defence can be non-paternalistically justified on the grounds that it is necessary for providing assurance to citizens that others are complying with their justice-based duties:

The state's coercive power may be necessary to provide the requisite assurance to each citizen that others will do their fair share. It seems plausible to suppose that even if we are all committed to acting justly, we cannot know that everyone else is similarly committed. Assuming that it is not irrational or unreasonable to refuse to do one's part in a cooperative scheme without the requisite assurance that others will do likewise, there thus need be no negative judgement implied when the state provides the requisite assurance by coercively enforcing certain duties of justice. (Quong 2011: 103)

The problem with this response, though, is that if we owe perfectionist duties to others, as duty-based perfectionists maintain, then perfectionists can likewise argue that public action is necessary for providing each citizen with the assurance that others are complying with their perfectionist duties. Just as we need assurance that others are discharging their duty to pay taxes towards neutral ends such as education, we also need assurance that others are discharging their duty to pay taxes towards perfectionist ends such as art galleries. Without such assurance, many perfectionist duties would plausibly be left unfulfilled. Perfectionists, too, that is, can argue with great plausibility that the state's coercive power may be necessary to provide the requisite assurance to each citizen that others will do their fair share towards the promotion of the conditions that generally conduce to moral, intellectual, and artistic excellence. So, assuming that refusing to do one's part in a cooperative scheme aimed at promoting human flourishing without the requisite assurance that others will do likewise is no more irrational than refusing to do one's part in a cooperative scheme aimed at promoting justice without such assurance,

there need be no negative judgement implied when the perfectionist state provides this assurance by coercively enforcing certain perfectionist duties.¹⁵

It is worth noting here that mutual assurance of duty compliance is not the only, or perhaps even the most important, non-paternalistic justification available to perfectionists. Indeed, the idea that the state functions as a mutual assurance mechanism is quite a narrow conception of the role of the state. There are, in addition, other non-paternalistic justifications for perfectionism, such as arguments that have to do with externalities, coordination problems, and public goods (see Quong 2011: 88–91; Kramer 2017: 63–91). My point here is simply that once we accept the duty-based perfectionist's claim that there are perfectionist duties, we can answer the "why is state action necessary?" question in the same way that neutralists do, namely by appealing to the role of the state in providing each citizen with assurance that others have complied with their duties.

3.3. The Quidnunc Mentality

In his recent book *Liberalism with Excellence*, Kramer develops an intriguing and original objection to perfectionism in its mainstream "edificatory" form—namely, that edificatory-perfectionist regimes are guilty of an objectionable "quidnunc" mentality, one akin in important respects to that of a curtain-twitching village busybody who distributes "boxes of delicious homemade fudge" to "inspire her fellow inhabitants

¹⁵ Of course, perfectionists could also pursue a more hardline response to the paternalism objection by saying (i) that there is *nothing* wrong with paternalism, or (ii) that the prima facie wrongness of paternalism is so uninterestingly *slight* that it would be routinely outweighed by the great benefits of perfectionist policies. For Quong (2011: 86), paternalism is prima facie wrong because it denigrates or disrespects a citizen's moral status by making "negative assumptions about citizens' abilities to know and rationally pursue their own good". But what if these assumptions are true, as psychological research increasingly suggests? Is it so denigrating or disrespectful to a person to proceed on the assumption—surely true of *all* human beings—that she will make rational mistakes about what is in her interests? To pose this question is not, as Quong (2011: 107) suggests, to call into question the "liberal conception of ourselves as free and equal". It is simply to call into question a particular *interpretation* of free and equal liberal citizenship—an interpretation that equates being treated as free and equal with being treated as if one were a superhumanly reliable practical reasoner.

to lead lives of propriety” and who in general meddles in matters that are “none of [her] business” (Kramer 2017: 251–86, at 282, 287). Thus whereas many critics of edificatory perfectionism adopt a citizen-focused perspective—contending, for instance, that perfectionism treats citizens infantilizingly or unequally—Kramer (2017: 254) instead adopts “a change of perspective” and focuses on the character and outlook of an edificatory-perfectionist system of governance and its state officials.

For Kramer (2017: 277), the quidnunc mentality is objectionable in two main ways, namely by being “a form of self-aggrandizement and a form of self-abasement”. It is meddlesomely *self-aggrandizing* because perfectionist policies “are aimed neither at preventing the infliction of harm by some people on others nor at fulfilling any of the further responsibilities” of government such as promoting public order (Kramer 2017: 277). And it is demeaningly *self-abasing* because “it presumes that the success of a system of governance is partly dependent on the willingness of citizens to alter their private pursuits in response to inducements by the system’s officials” (Kramer 2017: 277). As he elsewhere puts it, the edificatory-perfectionist state is self-abasing because it makes its success “dependent on choices *that are none of its business*”—choices that are private, choices about which it should not care (Kramer 2017: 298). An edificatory-perfectionist state thus violates a “deontological ethic of self-restraint”—a wrong that “tarnishes its moral integrity”, that “far-reachingly informs and taints” its relationship with the citizenry, and that “deprives [its edificatory-perfectionist laws] of moral legitimacy” (Kramer 2017: 253, 277, 278).

This argument has no force, however, against the duty-based perfectionist, who holds that governments have duties of edification towards their citizens. For duty-based perfectionists, the choices that citizens make about how to live—for instance, the choice to waste away one’s life in front of the TV—are not “private pursuits” but are instead a legitimate concern of the state. The state might, for instance, have duties to steer people away from a life wasted away in front of the TV and towards more worthwhile and meaningful avenues. In such a case, an edificatory-perfectionist state is no more self-abasing or self-aggrandizing than a neutral state that pursues imperatives such as justice and security, because their edificatory-perfectionist policies are in fact “necessary for the fulfillment

of the elementary moral responsibilities that are incumbent on any system of government” (Kramer 2017: 285). Another way of putting this point is as follows: when Kramer says that edificatory-perfectionist states stick their nose where it does not belong, he already assumes that states have no duties to promote the edification of the citizenry; but he cannot help himself to this assumption because whether the edificatory impulse falls within or outside the proper bounds of government is precisely what is at stake here. As before, then, the whole cogency of the quidnunc mentality objection rests on the presupposition that perfectionist duties do not exist—a presupposition that is explicitly rejected by duty-based perfectionism.¹⁶

4. Does Duty-Based Perfectionism Violate Public Justification?

Let me close by defending duty-based perfectionism against an objection that has traditionally been levelled against perfectionist theories, and that would seem to tell even more forcefully against the stronger duty-based form of perfectionism developed here. The objection is that perfectionism runs afoul of the public justification principle—the requirement, that is, that political decisions be justifiable to all reasonable citizens. As political liberals and public reason liberals frequently remark, perfectionist ideals of the good life and of human flourishing are controversial. Sincere, good-willed, intelligent people disagree about what constitutes human flourishing. So wouldn’t duty-based perfectionism involve the promotion of ends that some citizens reasonably reject? Couldn’t the claim that states have a duty to subsidize art galleries, for instance, be reasonably rejected by the average Joe who prefers to stay at home and watch TV? Taking seriously the fact of reasonable pluralism within modern democratic societies, would efforts by the state to discharge its perfectionist duties survive the public justification test?

¹⁶ It is worth noting that Kramer (2017: 289–96) is aware of a reply of this kind, but he rejects it largely on the same grounds as Quong, namely that duty-based edificatory perfectionism would be indistinguishable from liberal neutrality. I consider the issue of indistinguishability in Section 2.4.

“The deeply objectionable characteristic” of perfectionist politics, says Gerald Gaus in a forceful rendition of this worry, is that it makes the terms of our common life “dictated by the conscience of some, demanding that others simply obey” (2010: 230). He continues:

Social interaction is to be on their terms, directed by their conscience and their convictions about moral truth. Those who do not share this conscience must submit to the truth. It is precisely this imperious claim of private conscience that the social contract-liberal tradition has rejected . . . Surely the great liberal lesson, first glimpsed in the sixteenth century, is that successful large-scale cooperative social orders are precisely those in which people do come to “bracket” their rationally controversial first-personal convictions about the truth when these convictions are about how others must live. (Gaus 2010: 230–2)

Broadly speaking, there are two ways that duty-based perfectionists can reply to this objection. The first and more hardline response is to simply reject the public justification principle. The public justification principle, after all, is not self-evident. Justifiability to all reasonable citizens is quite a demanding requirement, and many remain unconvinced that public justification is a genuine constraint on government action. At most, say the critics, acceptability-to-all is a *pro tanto* value: it would be nice, of course, if all reasonable citizens could accept all laws, but this kind of acceptability is just another item on the long and dense list of political desiderata, and not a particularly weighty item at that.¹⁷ So duty-based perfectionists can just throw the ball back into the objector’s court by pointing out that public reason liberals need to provide a justification (or grounding, or moral foundation) for public justification. Public reason liberals need, that is, to answer the question: Why reason publicly? Why care about justifiability to all reasonable citizens? Why not simply implement the political rules and institutions that are *correct*—or at least that are most likely to be correct—regardless of whether those rules and institutions could be reasonably rejected by some citizens?

¹⁷ See, e.g., Enoch (2015: esp. 138–40) and Wendt (2019).

Of course, defenders of the public justification principle have sought to provide answers to these important questions. Charles Larmore, for instance, argues that public justification is required by respect for persons: “if we try to bring about conformity to a rule of conduct solely by the threat of force [and not by offering reasons that others recognize as valid], we shall be treating persons merely as a means, as objects of coercion, and not also as ends, engaging directly their distinctive capacity as persons” (1999: 607).¹⁸ But each of these attempts has been subjected to powerful critiques. For instance, while Larmore is surely correct that *some* failures to deliberate in terms of public reasons and to make publicly justified decisions would indeed be disrespectful (e.g. dictatorship or theocracy), it is unclear why *all* alternative models of democratic deliberation and decision-making short of public reason liberalism are incompatible with respect for persons. Even if a dictatorship treats persons as mere objects of coercion, why does a deliberative-democratic approach which requires citizens to favour only policies for which they have a high degree of rational justification, to deliberate openly and honestly, and to refrain from supporting laws that violate the dignity of others—but which does not require citizens to withhold support for policies that are not publicly justified—treat citizens disrespectfully and as mere objects of coercion?¹⁹ Needless to say, public reason liberals cannot answer this question by saying that deliberative democracy is disrespectful because it permits states to advance laws that some citizens reasonably reject—for that would amount to a bare assertion that respect requires public justification, and would beg the question against deliberative democrats for whom adherence to the norms of deliberative democracy is perfectly adequate to express respect for persons.²⁰

¹⁸ For other renditions of the argument from respect to public justification, see, e.g., Boettcher (2007) and Nussbaum (2011).

¹⁹ See, e.g., Eberle (2002: 84–151).

²⁰ Moreover, even if it could be shown that the failure to publicly justify is disrespectful *in some sense*, this does not look like an especially grievous kind of disrespect. As Lott puts it (2006: 87), “respect comes in shades”: it is disrespectful to cut in front of someone in a queue; it is even more disrespectful to spit on someone who supports a different football team; and it is yet more disrespectful to maim and torture someone on the basis of race or religion. The kind of disrespect involved in the failure to reason publicly looks fairly slight. Public reason, that is, looks to be at the periphery of our concept of respect rather than a paradigmatic instance of disrespect. See Lott (2006: 87–9, 95). This worry about weightiness becomes particularly acute once we observe that there are other important senses or dimensions of respect that are in

The second and more conciliatory response is to accept that public justification is a genuine constraint on governmental action but to argue that a good amount of duty-based perfectionism would survive the public justification test. This is the response of so-called moderate perfectionists, and is the response that I myself favour and have sought to defend in other work (Tahzib 2019). Duty-based perfectionists, that is, can follow Rawls in taking a “political turn” and in presenting their view in a way that is “political not metaphysical”. A “political” form of duty-based perfectionism is one that combines the contractualist commitment to public justification favoured by political liberals with the duty-based-perfectionist claim that the state must promote the good life. By taking a political turn of this kind, duty-based perfectionists can incorporate many of the central insights of political liberalism—that political philosophers must show greater sensitivity to reasonable pluralism, and that the terms of our common political life should as far as possible be justifiable to all—whilst also avoiding its rather less appealing requirement that we “cordon off political morality from our best understanding of human flourishing” (Wall 1998: 123). On this approach, then, certain perfectionist commitments, such as moral, intellectual, and artistic excellence, are built into the notion of reasonableness, in much the same way that many political liberals make reasonableness conditional on acceptance of certain liberal commitments such as freedom, fairness, and equality. Duty-based-perfectionist laws and policies such as state subsidies for art galleries would thus be expected to pass the test of public justification because a citizen who rejects the value of artistic excellence would count as unreasonable and so her veto would have no more normative force than that of a citizen who rejects the value of equality.

tension with the dimension of respect picked out by public reason liberals. For Galston, for instance, “we show others respect when we offer them, as explanation, what we take to be our true and best reasons for acting as we do” (1991: 109). See also Raz (1986: 157). So, in order to ground the kind of robust public justification principle that they are after—a principle of “very great and normally overriding weight”, in Rawls’s words (2005: 241)—public reason liberals need to explain why we should care so much about the specific kind of respect (or disrespect) expressed in giving (or failing to give) public reasons. Why does this highly specific and seemingly peripheral kind of respect generally take priority over (a) other senses or dimensions of respect and (b) other values in general?

The vital question that arises here is: Is this the right justificatory constituency? What justifies the claim that reasonableness is conditional not only on acceptance of liberal values but on acceptance of perfectionist values too? In particular, should a person who accepts freedom, equality, and fairness, but who rejects the value of artistic excellence, really be excluded as “unreasonable”? Is such a person really beyond the justificatory pale, in the way that a racist who rejects the value of equality is beyond the justificatory pale? Without a clear answer to these questions, the strategy of baking perfectionist content into the constituency of reasonable citizens looks unmotivated and gerrymandered—and thus unconvincing in comparison to Rawls’s and Quong’s more deeply theorized neutralist accounts of reasonableness.

It is not possible to address this concern fully here. But here is the sketch of a reply. For Rawls (2005: 9), the bounds of reasonableness are set by “a fundamental organizing idea within which all ideas and principles can be systematically connected and related. This organizing idea is of society as a fair system of social cooperation between free and equal persons”. I agree with this way of supplying content to reasonableness. But I am doubtful that the political liberal conception of society—a fair system of social cooperation between free and equal persons—is the only, let alone the best, such conception available. In particular, in other work I argue that a plausible and arguably superior alternative conception of society is that of society as a fair striving for human flourishing between free and equal persons. The exclusion of the art-hater from the constituency of reasonable citizens, then, is not an ad hoc fix to ensure that perfectionist policy achieves public justification but rather flows from deeper ideas about the nature of society which set the bounds of reasonableness.²¹ The art-hater is excluded because she fails to honour the idea of society as a striving for human flourishing. In other words, while the exclusion of the art-hater may be stipulative, it is not *arbitrarily* stipulative: these stipulations are well-motivated and driven by other theoretical

²¹ Naturally, this move raises further questions and worries. One such objection is: Doesn’t this make the public justification machinery redundant? Isn’t all the fundamental moral work being done by the perfectionist conception of society that sets the bounds of reasonableness? I address this worry in Tahzib (2019).

commitments; there are deeper reasons for making these particular stipulations rather than others.

Of course, there are limits to how far this strategy can go whilst remaining faithful to the basic public reason motivations.²² It would clearly make a mockery of the public justification apparatus if we held a view such as “Christianity: Political not Metaphysical”, according to which reasonableness is conditional on acceptance of Christian doctrine, where this conception of the constituency of reasonable citizens is in turn justified by appeal to a Christian conception of society. A view that defines “reasonable disagreement” in a way that bears no relation at all to the folk concept of reasonable disagreement is not aptly characterized as a “public reason” view. But I see no reason to think that these limits of fidelity are necessarily coextensive with liberal neutrality. A moderately perfectionist view of the kind that I propose—according to which members of the justificatory constituency endorse both liberal axioms (such as freedom, equality, and fairness) and perfectionist axioms (such as moral, intellectual, and artistic excellence)—would still leave room for the fact of reasonable disagreement to play a role (such as with respect to religious matters), albeit a more circumscribed role than it plays within political liberalism, and so such a view could plausibly be regarded as true to the spirit of the public reason project and faithful to its underlying motivations. This kind of perfectionism, I hope to have shown, can thus occupy a sweet spot by being thin enough that it satisfies a plausible specification of the public justification condition, yet thick enough that its conception of fair terms is distinguishable from the views of leading liberal neutralists such as Dworkin, Rawls, and Quong.²³

²² I thank Tom Sinclair for encouraging me to clarify this point.

²³ For very helpful comments and discussion, I thank Raven Adams, Max Afnan, Simon Căbulea May, Jeff McMahan, David Miller, Tom Sinclair, Steven Wall, and an anonymous reviewer. I also owe thanks to the participants of the 7th Annual Workshop for Oxford Studies in Political Philosophy, at Syracuse University, at which I presented this paper in August 2019, as well as to the participants of a departmental colloquium at the University of Southern California, at which I presented this paper in January 2020. I am especially grateful to Jonathan Quong, who has endured this paper on three occasions (including as my respondent in Syracuse) and who provided me with immensely generous and illuminating feedback.

References

- Arneson, R., "Paternalism, Utility, and Fairness", *Revue Internationale de Philosophie* 43 (1989), pp. 409–37.
- Boettcher, J., "Respect, Recognition, and Public Reason", *Social Theory and Practice* 33 (2007), pp. 223–49.
- Caney, S., "Liberal Legitimacy, Reasonable Disagreement, and Justice", *Critical Review of International Social and Political Philosophy* 1 (1998), pp. 19–36.
- Cavanagh, M., *Against Equality of Opportunity* (Oxford: Clarendon Press, 2002).
- Chan, J., "Legitimacy, Unanimity, and Perfectionism", *Philosophy & Public Affairs* 29 (2000), pp. 5–42.
- Couto, A., *Liberal Perfectionism* (Berlin: De Gruyter, 2014).
- Eberle, C., *Religious Conviction in Liberal Politics* (Cambridge: Cambridge University Press, 2002).
- Enoch, D., "Against Public Reason", in D. Sobel, P. Vallentyne, and S. Wall (eds), *Oxford Studies in Political Philosophy, Volume 1* (Oxford: Oxford University Press, 2015), pp. 112–44.
- Galston, W., *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State* (Cambridge: Cambridge University Press, 1991).
- Gaus, G., *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World* (Cambridge: Cambridge University Press, 2010).
- Kramer, M., *The Ethics of Capital Punishment: A Philosophical Investigation of Evil and Its Consequences* (Oxford: Oxford University Press, 2011).
- Kramer, M., *Liberalism with Excellence* (Oxford: Oxford University Press, 2017).
- Larmore, C., *Patterns of Moral Complexity* (Cambridge: Cambridge University Press, 1987).
- Larmore, C., "The Moral Basis of Political Liberalism", *The Journal of Philosophy* 96 (1999), pp. 599–625.
- Lott, M., "Restraint on Reasons and Reasons for Restraint: A Problem for Rawls' Ideal of Public Reason", *Pacific Philosophical Quarterly* 87 (2006), pp. 75–95.

- Metz, T., "Respect for Persons and Perfectionist Politics", *Philosophy & Public Affairs* 30 (2001), pp. 417–42.
- Nussbaum, M., "Perfectionist Liberalism and Political Liberalism", *Philosophy & Public Affairs* 39 (2011), pp. 3–45.
- Patten, A., "Liberal Neutrality: A Reinterpretation and Defense", *The Journal of Political Philosophy* 20 (2012), pp. 249–72.
- Quong, J., *Liberalism Without Perfection* (Oxford: Oxford University Press, 2011).
- Quong, J., "Liberalism Without Perfection: A Précis", *Philosophy and Public Issues* 2 (2012), pp. 1–6.
- Rawls, J., *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).
- Rawls, J., *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001).
- Rawls, J., *Political Liberalism: Expanded Edition* (New York: Columbia University Press, 2005).
- Raz, J., *The Morality of Freedom* (Oxford: Oxford University Press, 1986).
- Richards, J., "Equality of Opportunity", in A. Mason (ed.), *Ideals of Equality* (Oxford: Blackwell, 1998), pp. 52–78.
- Sandel, M., *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press, 1998).
- Schouten, G., *Liberalism, Neutrality, and the Gendered Division of Labor* (Oxford: Oxford University Press, 2019).
- Sher, G., *Beyond Neutrality: Perfectionism and Politics* (Cambridge: Cambridge University Press, 1997).
- Sher, G., "Confessions of a Quidnunc", *The American Journal of Jurisprudence* 63 (2018), pp. 49–61.
- Tahzib, C., "Perfectionism: Political Not Metaphysical", *Philosophy & Public Affairs* 47 (2019), pp. 144–78.
- Tahzib, C., *A Perfectionist Theory of Justice* (unpublished manuscript).
- Wall, S., *Liberalism, Perfectionism and Restraint* (Cambridge: Cambridge University Press, 1998).
- Wall, S., "Enforcing Morality", *Criminal Law and Philosophy* 7 (2013a), pp. 455–71.

Wall, S., "Moral Environmentalism", in C. Coons and M. Weber (eds), *Paternalism: Theory and Practice* (Cambridge: Cambridge University Press, 2013b), pp. 93–114.

Wendt, F., "Rescuing Public Justification from Public Reason Liberalism", in D. Sobel, P. Vallentyne, and S. Wall (eds), *Oxford Studies in Political Philosophy, Volume 5* (Oxford: Oxford University Press, 2019), pp. 39–64.

When Public Reason Falls Silent

Liberal Democratic Justification versus the Administrative State

Brian Kogelmann and Stephen G. W. Stich

[The traditional parameter of elasticity] cannot be used directly in the type of model developed by the Congressional Budget Office (CBO) because a percentage increase in quantity cannot be applied to a firm that initially does not offer insurance in the base case, where the initial offer rate is zero. (The same logic applies to individual take-up of insurance coverage.) To derive the likelihood of a new offer for a currently nonoffering firm, the CBO model employs a converted elasticity, called a take-up elasticity:

$$\varepsilon_t = \varepsilon x \frac{Q_1}{1 - Q_1}$$

where ε is the literature-derived offer elasticity and ε_t is the take-up elasticity.¹

§1. Introduction

The authors do not understand what the above quotation means. Yet this is arguably a good thing. The epigraph discusses the United States

¹ Congressional Budget Office, *CBO's Health Insurance Simulation Model: A Technical Description* (2007): 29, <http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/87xx/doc8712/10-31-healthinsurmodel.pdf>.

Congressional Budget Office's (CBO's) method of interpreting the elasticity of demand for health insurance. The Department of Health and Human Services (HHS) relied on the CBO's assumptions when estimating the impact of health insurance exchanges in its regulatory impact analysis accompanying proposed rules to implement portions of the Patient Protection and Affordable Care Act (PPACA), one of the most significant pieces of American legislation in the last few decades. Given the importance and impact of the rules implementing the PPACA, it was imperative for HHS to use sophisticated methods of analysis that the authors of this chapter—neither economists nor specialists in healthcare—lack the background to comprehend. The thesis of this chapter is that this need for highly sophisticated analysis by administrative agencies is incompatible with the most prominent account of liberal democratic justification in the current moral and political philosophy literature: John Rawls's idea of public reason. In short: public reason is inconsistent with the administrative state.

Our basic argument runs as follows. Broadly speaking, the demands of public reason require citizens to justify policies on the basis of reasons all citizens accept or endorse. These reasons include scientific and social scientific considerations. Thus, insofar as these considerations are employed to justify policies, they must be such that all citizens accept or endorse them. However, given (i) reasonable disagreement among professional scientists and social scientists over which theories are correct, most appropriate, most fecund, etc., and (ii) citizens' lack of competent training and thus lack of complex scientific knowledge, the set of scientific reasons that citizens may appeal to in public debate is astonishingly small. Public reason falls silent.

Now, one benefit of having administrative agencies is that it puts complex decisions in the hands of capable experts. What does the average citizen or member of Congress know about health insurance demand elasticity? Not much, but this is not a problem if they are not the ones setting the nitty-gritty of health insurance exchange rules. Better this be done by the experts at HHS. But if the average citizen knows nothing about health insurance demand elasticity, then the demands of public reason prohibit anyone—including the experts at HHS—from justifying health insurance exchange rules by appealing to facts about

health insurance demand elasticity, as they do in the epigraph to this chapter. Public reason is silent on this issue, so the demands of public reason prevent health insurance experts from appealing to their expertise when deciding the appropriate regulation. More generally, there is a direct tension between the requirements of public reason and one important and salutary fact about the administrative state. Public reason requires citizens to appeal to reasons and thus scientific and social scientific considerations that all can accept or endorse, but the administrative state is desirable *precisely because such a set of reasons does not exist*.

The problems for public reason do not end here. Following Rawls, most contemporary proponents of public reason favor a fairly robust version of egalitarianism. But material equality is achieved largely through decisions made by administrative agencies that can be justified only on the basis of non-public reasons. Health insurance exchange rules are a prime example. Their purpose is to make healthcare more accessible to the least advantaged, but they are justified on the basis of complex facts about health insurance demand elasticity and the like, which fall outside the scope of public reason. The upshot? The methods of achieving an egalitarian distribution that proponents of public reason often advocate are inconsistent with their more general commitment to the idea of public reason as a theory of liberal democratic justification. Indeed, we are skeptical that there is any way of achieving an egalitarian distribution that satisfies the demands of public reason.

The structure of the chapter is as follows. In §2 we briefly introduce the idea of public reason. In §3 we argue that the demands of public reason apply to scientific and social scientific reasons when used to justify policies. In §4 we argue that public reason implies that the administrative state is impermissible, and that its demands prevent the administrative state from being even marginally effective. In §5 we demonstrate that public reason condemns the tools required for achieving an egalitarian distribution. In §6 we entertain and refute four objections to our central argument: that publicly justifying a constitution, or perhaps the institution of delegation, which does not require expert scientific and social scientific knowledge suffices to justify administrative policies; that idealization can circumnavigate disagreement and ignorance on scientific and

social scientific matters; that citizens can come to learn the relevant scientific and social scientific reasons through engagement with expert public intellectuals; and that public reason is concerned with legitimizing only the ends of policy, whereas the administrative state deals strictly with the means of policy implementation. There is a concluding section.

§2. The Idea of Public Reason

Public reason approaches to justification begin with the fact of reasonable pluralism. As Rawls describes it: “A modern democratic society is characterized not simply by a pluralism of comprehensive religious, philosophical, and moral doctrines, but by a pluralism of incompatible yet reasonable comprehensive doctrines” (Rawls 2005: xvi). Two things are worth noting about the fact of reasonable pluralism. First, reasonable pluralism does not occur by mere chance. Rather, we should *expect* reasonable pluralism to obtain in a society governed by liberal institutions, for reasonable pluralism results from persons exercising the freedom of thought and expression: “a plurality of reasonable yet incompatible comprehensive doctrines is the normal result of the exercise of human reason within the framework of the free institutions of a constitutional democratic regime” (Rawls 2005: xvi).

Second, the fact of reasonable pluralism is not identical to the fact of pluralism, or disagreement more generally. There are many kinds of disputes in which persons find themselves, but not all such disagreements are reasonable and thus instances of reasonable pluralism. As Rawls notes, disagreement may occur because “people hold views to advance their own more narrow interests,” or “perhaps people are often irrational and not very bright, and this mixed with logical errors leads to conflicting opinions” (Rawls 2005: 55). The fact of reasonable pluralism is not meant to encapsulate these sorts of disagreements. Rather, reasonable pluralism refers to those disputes that are the result of the *burdens of judgment*, which are a set of factors that lead to disagreement. If disagreement is the result of these factors, then the disagreement is reasonable; if disagreement is *not* the result of these factors, then the disagreement is not reasonable. To offer a brief summary of the burdens of judgment:

(a) evidence is often conflicting and complex, and thus hard to assess and evaluate; (b) we disagree over how much weight to assign relevant considerations; (c) concepts are vague and subject to hard cases, requiring interpretation; (d) the way we assess evidence and moral and political values is shaped by our total life experience; (e) there are often different kinds of considerations of different forces on both sides of an issue; and (f) social systems are limited in the values they can realize, and some selection must be made (Rawls 2005: 56–57). When any one of (a)–(f) leads to disagreement, then we say that such disagreement is an instance of reasonable pluralism.

In the face of reasonable pluralism, public reason approaches to justification demand that those engaged in justification avoid controversial issues, and that when justifying political proposals to one another citizens rely on considerations that all can accept or endorse. As Jonathan Quong describes it in his excellent survey article: “The idea of public reason . . . is a view about what kinds of reasons citizens in a well-ordered democratic society ought to invoke when deciding important political questions. It asks us to refrain from appealing purely to religious or comprehensive doctrines over which reasonable people disagree, and instead to seek shared reasons acceptable to similarly motivated persons to justify political principles and laws” (Quong 2014: 270). For example, suppose Rowena is a Christian and Colin a Muslim. If Rowena wants to justify some policy *p* to Colin, then public reason demands that she refrain from relying on the Bible or other Christian doctrine, for these are subject to reasonable disagreement. She cannot argue for *p* by showing how certain parts of Christian scripture support *p*. Rather, public reason demands that Rowena appeal to considerations she and Colin share: perhaps that all citizens are born free and equal, and policy *p* will help better secure equality. To put it succinctly: “people should respond to points of disagreement by retreating to neutral ground, to the beliefs they still share” (Larmore 1996: 135).

Public reason is thus a constraint on the reasons citizens may give one another in public discourse.² Why accept such onerous restrictions? The

² A related research program says that the *rules* persons follow—like coercive laws or social-moral rules—must be in some sense acceptable to all who follow them. This approach is taken

general idea is that the very concept of justification *requires* us to rely on public reasons in our discourse with one another. As Rawls notes: “justification is addressed to others who disagree with us, and there it must always proceed from some consensus, that is, from premises that we and others publicly recognize as true” (Rawls 1999: 394). To give an example, suppose Rowena wants to justify some policy *p* to Colin, who is skeptical of it. For her to actually justify *p* to Colin, she must (according to Rawls) rely on premises that both she and Colin accept. But in doing so, Rowena ends up adhering to the demands of public reason, which say that when confronted with reasonable disagreement citizens must rely on shared considerations in their discourse. Hence, justification in the face of reasonable disagreement requires a reliance on public reason.³

To end, it is worth noting that there is some disagreement among its supporters concerning the *scope* of public reason. In particular: for what sorts of political questions must public reason be employed as the method of justification? Rawls’s view on this issue remains unclear. He says that public reason need only be employed when “matters of basic justice” and questions concerning “constitutional essentials” are at issue in the public sphere (Rawls 2005: 214). Sometimes he seems to think public reason should never be employed beyond this narrow set of issues (e.g., Rawls 2001: 91n13). In other places he suggests that, ideally, public reason should apply to all political questions (e.g., Rawls 2005: 215). More recently, Quong has argued that public reason should apply to all political issues involving coercion (Quong 2004). On this view, if policy *p* will be carried out by some coercive means, then *p* must be justified with shared reasons. This is the understanding of public reason’s scope that we shall adopt for the purposes of this chapter. Since our thesis in part

by the likes of Gaus (2011); Vallier (2019). Our concern in this chapter is with the idea of public reason, which is about the *reasons* persons may give one another in public justification. We are not sure what, if any, relevance our thesis has for the other research program that takes as its focus acceptability of the rules that persons follow. For more on the distinction between these two closely related research agendas, see Lister (2013: 15).

³ One might further wonder why we owe persons justification at all, or why we only owe persons justification when our disagreements with them are reasonable. This gets into the question of what *grounds* public reason (be it respect for persons, civic friendship, etc.), and why we should differentiate between the reasonable and unreasonable when it comes to our duties to others. Such issues are important, but beyond the scope of the current chapter. We assume that justification is owed to (at least) persons who reasonably disagree with us. This is sufficient to trigger the public reason requirement.

hinges on this wider account of public reason's scope, we shall offer a defense of it in §6.1 below, when we entertain objections to our thesis.

§3. Public Reason and the Sciences

It is clear from both the introduction to this chapter and simple reflection that many political questions cannot be settled by appeal to normative reasons alone. Debate over such questions will thus not only involve normative considerations—freedom, equality, and the like—but also positive considerations: considerations from the sciences and social sciences needed to adjudicate between different policy options that confront a polity. It is clear public reason demands that citizens only appeal to shared normative reasons when engaged in public discourse. Rowena the Christian cannot appeal to Christian moral teachings in her debate with Colin the Muslim but must appeal to considerations the two share: for instance, a common understanding of equality. It is less clear, though, whether this demand applies to scientific and social scientific considerations used in public discourse. Rawls seemed to think so. He tells us that, along with normative considerations, public reason contains “guidelines of inquiry: principles of reasoning and rules of evidence in the light of which citizens are to decide whether substantive principles properly apply and to identify laws and policies that best satisfy them” (Rawls 2005: 224). While Rawls's position seems clear, we should not just take him on authority. If Colin argues for policy *p* by appealing to social scientific considerations that Rowena does not accept, has he violated the demands of public reason?⁴

We believe that Colin *has* violated the demands of public reason. This is because many disagreements over scientific and social scientific questions result from the burdens of judgment, and are thus reasonable

⁴ The authors can find only two papers related to the topic of science and public reason. In the first, Jøneh-Clausen and Kappel (2016) examine what they take to be Rawls's account of when scientific considerations should be included in public reason. In the second, Badano and Bonotti (2020) examine whether the *accessibility theory of public justification*—a variation of Rawls's theory of public reason (which is what our chapter is concerned with)—would permit scientific considerations in public debate (to which they answer: yes).

disagreements. Since the public reason theorist adopts a general imperative to move to shared ground when confronted with reasonable disagreements, consistency demands that she also insist on shared science and social science in public debate when reasonable disagreement obtains in these domains. If the disagreement between Rowena the Christian and Colin the Muslim is reasonable, then the two must find shared ground according to public reason. Similarly, if the disagreement between Rowena the neoclassical economist and Colin the Austrian economist is a reasonable one, then by the same reasoning so too must the pair retreat to shared ground. Hence, public reason applies to scientific and social scientific considerations used in public discourse. What remains to be seen in more detail is whether scientific and social scientific disputes do in fact result from the burdens of judgment.

Let us examine this more carefully. In doing so, we shall focus mostly on disputes among economists, for this is where our familiarity lies. Consider one of the burdens of judgment, that disagreement can occur because “evidence—empirical and scientific—bearing on the case is conflicting and complex, and thus hard to assess and evaluate” (Rawls 2005: 56). Disagreement over how to assess evidence in the sciences is clearly illustrated by a recent event in the economics profession. The 2019 Nobel Memorial Prize in Economic Sciences was awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer for their work in development economics. These scholars’ main contribution is their innovative use of field experiments. The general idea is that we can learn about “what works” in development economics by testing theoretical proposals on a smaller scale. For instance: given that important public health professionals like nurses often do not show up for work in developing countries, would a system of monitoring and punishment decrease absences? The answer is no, as local health administrators in India responded to such an incentive system by increasing the number of permitted sick days for nurses (Banerjee *et al.* 2010).

Or, do microloans actually help reduce poverty? The results are ambiguous—investment in small business increases, but there was no meaningful change in health outcomes, education, or women’s empowerment (Banerjee *et al.* 2015). While impressive in their scope, several notable figures (including other Nobel Laureates) question how

much we can really learn from the evidence collected in these sorts of experiments (e.g., Deaton and Cartwright 2018). In a world characterized by complexity, how much can we really learn about the practice of microfinancing generally speaking by offering small loans to fifty-two randomly selected neighborhoods in Hyderabad, India? This is something economists disagree much about, and seems to be an example of one of Rawls's burdens of judgment in action: in particular, disagreement over how to assess the relevant evidence.

Another burden of judgment occurs when we "agree fully about the kinds of considerations that are relevant," but nonetheless "disagree about their weight, and so arrive at different judgments" (Rawls 2005: 56). Scientists and social scientists frequently disagree over how to weigh shared standards. Indeed, this is a point emphasized by philosopher of science Thomas Kuhn. Kuhn argued that scientists use five desiderata to appraise theories: accuracy, logical consistency, scope, simplicity, and fecundity (Kuhn 1977: 321–322; 2012: 184–185). These criteria, though, often conflict with one another. More accuracy will often yield less simplicity, and vice versa. Thus, scientists must make judgments about which of our conflicting desiderata to emphasize and which to discount—in other words, they must decide how to weigh the criteria. But which criteria to emphasize and which to discount is something scientists disagree over: "When scientists must choose between competing theories, two men fully committed to the same list of criteria for choice may nevertheless reach different conclusions...perhaps they agree about these matters but differ about the relative weights accorded to these or other criteria" (Kuhn 1977: 324). As an example of this, Colin might endorse theory T_1 because it does better across accuracy than theory T_2 ; Rowena might endorse T_2 because it does better across simplicity than T_1 . Such a dispute is clearly a result of Rawls's burdens of judgment.

A third burden occurs when our concepts or theories "are vague and subject to hard cases." Because this is so, "we must rely on interpretation (and on judgments about interpretation) within some range (not sharply specifiable) where reasonable persons may differ" (Rawls 2005: 56). Problems of interpretation and the disagreements that ensue are also frequently present in the sciences. Consider just one example, articulated

by economist Peter J. Boettke (1997). One of the most important technical results of twentieth-century economics is the so-called first fundamental theorem of welfare economics, proved by Kenneth J. Arrow and Gérard Debreu (1954). Roughly, this result shows that markets—when modeled under highly unrealistic assumptions—are always Pareto efficient: there exists no transfer of goods that could make some party better off without making at least one other party worse off.

Economists differ greatly over how to interpret the stunning theorem. The so-called Neo-Keynesians use the model “as a critical standard with which reality could be indicted when it failed to measure up” (Boettke 1997: 23). On this view, because the fundamental theorem depends on unrealistic assumptions, it suggests that markets in our actual world do *not* live up to the theorem’s optimistic results, and thus need intervention to correct their failures. On the flipside, Chicago school economists interpret the result as descriptively meaningful: “In their view, real markets come breathtakingly close to approximating the efficiency properties of general competitive equilibrium” (Boettke 1997: 23). Finally, an eclectic group of thinkers led by Ronald Coase interpret these models as “foils,” where “the descriptive value of the model lay precisely in its departure from observed reality, for this underscores the function of real-world institutions in dealing with *imperfect* knowledge, uncertainty, and so forth” (Boettke 1997: 23). So here we have three major schools of economic thought—all harboring numerous Nobel Laureates—whose disagreements are grounded in fundamental interpretive questions. This, surely, counts as an instance of the burdens of judgment in action.

More examples are available, but we do not want to belabor the point. The above case studies are sufficient to show that disputes in the sciences and social sciences often result from the burdens of judgment. Such disputes are instances of reasonable pluralism. Given this conclusion, public reason then demands that, when debating political questions in the public sphere, citizens must strictly rely on *shared* scientific and social scientific considerations in public debate, just as they must rely on shared moral considerations. Reasonable disagreement over normative considerations forces a retreat to shared ground, as does reasonable disagreement over scientific questions.

There are two general ways citizens can fail to live up to the demands of public reason when using scientific and social scientific considerations in public debate, and it will be helpful to distinguish them. In the first, called *controversy failure*, Rowena relies on scientific and social scientific consideration that Colin rejects. Here is an example. Rowena might look at a market with large information asymmetries and suggest that, because the assumptions of the first fundamental theorem don't hold, the market must be failing and requires intervention. Colin, though, rejects this interpretation of the theorem and, as a Chicago school economist, thinks that the market Rowena points to approaches the results of the theorem, and is thus efficient. No intervention is required, and would likely make things worse. Because Colin rejects the considerations Rowena relies on, she has failed to rely on shared reasons, and thus violates public reason.

The second way public reason can misfire when it comes to scientific and social scientific considerations is through *ignorance failure*, which occurs when Rowena relies on considerations that Colin does not understand. To give an example: Rowena attempts to argue in defense of policy p by appealing to a complex partial equilibrium model of a market where firms and consumers have asymmetric information sets, which Colin does not and moreover cannot understand. What Rowena says to him might as well be in another language. Unable to comprehend the social science Rowena relies upon, it can hardly be said that Colin accepts or endorses such considerations. Once again, Rowena has failed to rely on shared social scientific reasons in her debate with Colin, and thus violates the demands of public reason.

We now arrive at the main point of this section: the scientific and social scientific considerations that successfully avoid controversy and ignorance failure will be quite simple and rudimentary. In other words: the set of scientific public reasons will be quite small. This is something Rawls both recognized and seemed quite comfortable with. He twice says that persons in public debate cannot appeal to "elaborate economic theories of general equilibrium and the like" (Rawls 2005: 225; 2001: 90). This is the case even though the standard Arrow-Debreu general equilibrium model is *the* foundation of neoclassical economics, taught in nearly all mid-level undergraduate microeconomics classes in universities across

the United States. If mid-level undergraduate economics is off the table, then certainly everything one would learn in an economics PhD course sequence is off limits as well so as to avoid ignorance and controversy failure. We now go on to show what implications this has for an essential—and, we would argue, ineliminable—feature of contemporary liberal democracies: the administrative state.

§4. The Administrative State

One of the hallmarks of modern administrative agencies is their expertise. Members of the bureaucracy have (often advanced) degrees in the relevant fields and belong to relevant professional associations. For example, at the time of writing, the director of the Office of the Actuary at HHS's Centers for Medicare and Medicaid Services is a mathematician who is a member of the American Academy of Actuaries. The same often goes for politically appointed agency heads. At the time of writing, heads of seven of the nine operating divisions within HHS that have a permanent politically appointed leader have an advanced degree in either science, medicine, or public health—some have multiple advanced degrees. The backgrounds of employees and heads of administrative agencies should give a strong indication of the kinds of reasoning that these agencies employ in their work. If agencies did not need to engage in advanced scientific and social scientific reasoning, they would have no need to hire individuals with highly specialized backgrounds directly related to the agency's work. Constituting agencies this way would seem very peculiar.

Moreover, agencies' outputs often reflect very complicated scientific and social scientific analyses. HHS's application of the CBO analysis quoted in the epigraph is typical of regulatory impact analyses, which US federal agencies must perform for all significant regulatory actions (Executive Order 12,866, §§ 3(f)(1), 6(a)(3)(C) (1993)). These impact analyses must include, among other things, cost-benefit analyses (which require agencies to engage in complex processes of quantifying costs and benefits) and estimates of the effectiveness of non-regulatory economic incentives to accomplish the purposes of the regulation (Executive Order

13,563, § 1(b) (2011)). These analyses, necessarily, are extremely complicated. And this is desirable: one would hope that regulations affecting hundreds of millions of lives and trillions of dollars would incorporate the most sophisticated analyses available.

Next, the agency-court relationship—created both legislatively and judicially—is founded on agencies’ comparative scientific and social scientific expertise. Section 706 of the Administrative Procedure Act governs judicial review of agency action, findings, and conclusions. It gives courts the authority to set aside agency actions, findings, and conclusions only if they are “arbitrary and capricious” or “unsupported by substantial evidence,” which are generally understood to be interchangeable.⁵ The Supreme Court’s gloss on this standard from the seminal case of *Motor Vehicle Manufacturers Association v. State Farm Mutual Insurance Co.* is instructive: “Normally an agency rule would be arbitrary and capricious if the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise” (*Motor Vehicle Manufacturers Association* 1983: 43). This articulation of the standard of review (known as “hard look” review) makes clear that the agency, not the court, is the body that employs the expert reasoning—indeed, the courts have prohibited themselves from substituting their judgment for that of the agency (*Citizens to Protect Overton Park* 1971: 416). Instead, courts most often employ hard look review to invalidate agency action when the agency has failed to provide any justification whatsoever (Wald 1996: 234). This is precisely the type of review one would expect when the reviewing body is incapable of understanding, to any substantial degree, the reasoning in which the body being reviewed typically engages.

Furthermore, when courts find that an agency did not perform the necessary analysis in promulgating a rule, the remedy is usually to give the agency another opportunity to justify it, not to invalidate it (Meazell 2011: 741). For instance, the *Motor Vehicle Manufacturers Association*

⁵ The statute also contains grounds for invalidating agency actions that are not relevant here.

court rejected a rule promulgated by the National Highway Traffic Safety Administration (NHTSA) that would have permitted cars to be constructed without passive restraints such as airbags and automatic seatbelts because “NHTSA apparently gave no consideration whatever to modifying the [motor vehicle safety standard] to require that airbag technology be utilized . . . There was no suggestion in the long rulemaking process that led to [the standard] that, if only one of these options were feasible, no passive restraint standard should be promulgated” (*Motor Vehicle Manufacturers Association* 1983: 46). But, instead of holding the rule invalid, the court remanded the case to the agency to evaluate certain issues that “rest[ed] within the expertise of NHTSA” (*Motor Vehicle Manufacturers Association* 1983: 53). If the courts believed they had the capacity to understand whether the agency’s policy was justified by the relevant scientific or social scientific considerations, one would expect them to simply hold that the rule was arbitrary and capricious and invalidate it, not give the agency another chance. For instance, if the *Motor Vehicle Manufacturers Association* court took itself to be capable of determining whether airbag technology actually was necessary given the relevant engineering and economic reasons, it could have invalidated the rule.⁶ Courts are thus aware of the expert reasoning of administrative agencies, which even they, as entities that interact with agencies on a regular basis, do not deign to perform.

The composition of agencies, their complex output, and their relationship with courts all strongly suggest that they do not adhere to public reason when selecting policies to implement. As noted above, it would be peculiar for agencies to hire individuals with specialized backgrounds but for these individuals not to use their specializations when formulating agency policy. We should expect the actuaries employed by HHS to use their specialized actuarial knowledge when formulating health insurance policy. But, as we argued in the previous section, the demands of public reason require that the scientific reasons used in public justification—including actuarial reasons—be shared among all those who will be subject to such policies. Yet it is extremely implausible that most ordinary

⁶ It also could have upheld the rule if it took itself to be capable of determining that the agency was right for the wrong reasons.

citizens without at least a post-secondary education in mathematics, economics, or actuarial science understand the relevant considerations. So HHS does not reason with considerations all citizens accept when justifying healthcare policy. The demands of public reason thus prohibit HHS from appealing to such considerations when justifying the relevant policy. But how could exceedingly complex health insurance policies like the rules implementing the PPACA be justified *except* by reference to such complex and inaccessible considerations? They could not. Given the paucity of scientific public reasons that exist to justify policies, there is no way to justify important policies like these rules using strictly public reasons. The silence of public reason blocks most health insurance policies. Conversely, if agencies take into account more than very simple reasons—as they must, given the complexity of the problems they address—then they will violate the requirements of public reason. It follows that such (highly desirable) policies will fail to be publicly justified.

Note that our argument here is neutral with respect to justifications for having an administrative state in the first place. In particular, it does not presuppose the strong claim, most often associated with New Deal administrative law scholars like James Landis (1938), that the administrative state is justified because it places decision-making power in the hands of capable and independent experts. This view has been criticized by administrative law scholars, and does not have many adherents today.⁷ Our argument rests on a much weaker claim, namely that *if* there is to be an administrative state making highly consequential and detailed regulations, *then* it should be staffed by experts employing expert reasoning. This is consistent with the New Deal view, and also with more recent “civic republican” justifications of the administrative state, such as those of Henry Richardson (2002) and Mark Seidenfeld (1992), that emphasize agencies’ roles in deliberative decision-making. Indeed, Richardson believes it is *uncontroversial* that agencies should have scientific and social scientific expertise (Richardson 2002: 224). To be inconsistent with the requirements of public reason, all that need be true is that administrative agencies use expert reasoning. As a matter of fact, this is the case. And, if agencies are to be effective, it ought to be.

⁷ For a good summary of some of the major criticisms, see Gifford (1984: 312–319).

Finally, the problems we have observed with the administrative state thus far reflect only ignorance failure—agencies employ complex reasoning in promulgating rules that nearly all persons subject to such rules do not understand. They thus do not accept or endorse them. Indeed, we believe this is the most significant problem for the administrative state because, for any given policy, the vast majority of persons are non-experts in the relevant fields. But controversy failure is also a serious worry. This is because complex science and social science are often just as controversial as complex philosophical positions, and agencies must choose which considerations to rely on when making policy. Inevitably, this will entail drawing on controversial consideration that, while perhaps *understood* by all experts in the relevant field, are not *accepted* by all of them.⁸ The requirements of public reason would then prohibit the agency from using those reasons to justify the policy *even if* the ignorance problem did not obtain (which it most certainly does). As Adrian Vermeule notes in his recent book on administrative law: “Increasingly, agencies operate at a policymaking frontier where settled science and expert consensus have run out. When agencies do face serious uncertainty of this sort, reasons themselves have run out as well” (Vermeule 2016: 126). Public reason, Vermeule agrees, falls silent.

§5. Egalitarianism versus Public Reason

Since public reason is inconsistent with the administrative state, it is also inconsistent with forms of political organization that require the administrative state. This point is obvious, but should be troubling for most proponents of public reason. This is because most proponents of public reason advocate some kind of egalitarian distribution, but achieving such a distribution requires an administrative state that we have just shown to be inconsistent with public reason. We do not claim that it is conceptually or logically impossible to achieve an egalitarian distribution without violating the demands of public reason. But we do claim that the

⁸ To see that this occurs, one need only read a small number of the more than 3,000 comments submitted by members of the public in the notice and comment process for the two proposed rules implicated by the HHS regulatory impact analysis quoted above.

demands of public reason prohibit every major Rawlsian-egalitarian proposal in the existing literature. Since it is impractical to evaluate every proposal here, we focus in on Rawls's proposal for achieving material equality in particular, and show why it violates the demands of public reason. We believe the argument generalizes, *mutatis mutandis*, to other institutional proposals for achieving an egalitarian distribution.

Every complete proposal for achieving material equality involves some combination of specific administrative components. While Rawlsians and egalitarians more generally differ on the exact combination of administrative components, they generally favor a large and diverse set of agencies to redistribute goods in different ways so as to achieve equality. Rawls is a paradigm example. In speculating about the institutions of a just society he proposes (1) the *allocation branch*, which ensures competitive prices and healthy market competition through taxes, subsidies, and the redefinition of property rights; (2) the *stabilization branch*, which focuses on keeping unemployment at a minimum and increasing aggregate demand; (3) the *transfer branch*, which sets and distributes the social minimum; and (4) the *distribution branch*, which imposes taxes, redistributes shares, and redefines property title in order to achieve justice (Rawls 1971: §42).

It should be easy to see that the administrative agencies Rawls proposes will employ methods of reasoning likely at odds with the demands of public reason. In fact, the demands of public reason likely prohibit *each and every branch* of Rawls's ideal bureaucratic state. Ordinary citizens do not share the economic considerations that would justify intervention in the market through taxes, subsidies, and redefinition of property rights in order to ensure competitive prices. This can be seen by simply doing a brief survey of the optimal taxation and public finance literature more generally (e.g., Tuomala 2016). Likewise, citizens do not know the models and measurement techniques that would be used to reduce unemployment while simultaneously increasing aggregate demand. The difficulties here can be seen from a quick glance at the literature on welfare economics and market failures (e.g., Salanie 2000), as well as by examining the debate surrounding Keynesian economic stimulus (e.g., Friedman 1957). These debates also show that experts disagree on the relevant social scientific considerations. Because of this, it

seems that both Rawls's allocation branch and stabilization branch will employ methods of reasoning at odds with the demands of public reason: such bureaucracies will not justify policies by relying on shared reasons. Faced with this fact, the Rawlsian must decide which is more important: justification through public reason, or an egalitarian distribution?

It might be suggested that Rawls's transfer branch runs into less trouble than the allocation and stabilization branches, because the central goal here is to simply set a social minimum. More generally, a universal basic income seems less objectionable from the point of view of public reason than other proposals because it does not depend, at least as much, on predicting the effects of large-scale regulation—there is simply much less regulation to think about, so fewer complex and controversial considerations will enter the justification of policy. One sets a baseline income for persons and then lets the market process take hold. While admittedly less objectionable, even so simple a distributive policy as a universal basic income will be at odds with public reason for several reasons. First, it will be necessary to model and measure the economic incentives—particularly in the labor market—of providing individuals with a universal basic income in order to determine the appropriate income to provide. This requires using complex economic models, statistical methods, and quantitative analysis, most of which will be drawn from non-public reasons. Second, it will be necessary to determine how to collect the wealth to be distributed, consistent with other important considerations like economic efficiency. This requires choosing from among many possible taxation rates of many different kinds of wealth (personal income, corporate income, consumption, gift, inheritance, etc.), which in turn requires understanding the interdependencies among different kinds of tax systems—again, something of which ordinary citizens have little to no grasp. There will also be interactive effects between the incentive question and the taxation question, further taking matters out of the intellectual capacities of most citizens. And all of this is undoubtedly controversial among experts. Public reason simply runs out here.

Finally, there is the distribution branch, whose function is to achieve a just distribution. By this point it should be clear that this branch will

employ methods of reasoning at odds with the demands of public reason. Take healthcare as an example. Norman Daniels argues that Rawlsian distributive justice requires some form of public healthcare, since this is necessary to achieve fair equality of opportunity. One of Daniels's suggestions is a tiered healthcare system, where the basic tier is financed by a "national health insurance scheme that eliminates financial barriers" (Daniels 1981: 176). But, as discussed above, a health insurance scheme relies on facts about health insurance demand elasticity and the like that are not widely shared. Other types of healthcare, such as single-payer systems, will depend on other complex and controversial financial, economic, and actuarial facts, not to mention unshared facts about medicine and public health. Since there are no public reasons that suffice to justify public healthcare schemes, the demands of public reason block the justification of public healthcare schemes. Of course, the point here generalizes: any kind of complex social engineering done in the name of distributive justice will be blocked by public reason, for such complex social engineering will require an administrative state that relies on complex scientific and social scientific considerations that many do not share.

There is thus good reason to think that the typical proposals for achieving egalitarianism are inconsistent with public reason because of their reliance on administrative agencies and thus expertise. Note, in arguing for this, we are not claiming that *only* egalitarianism is inconsistent with public reason. We merely focus on egalitarianism because it is the substantive theory of distributive justice that most defenders of public reason happen to endorse. Even so-called classical liberalism is probably inconsistent with public reason, as classical liberals such as F.A. Hayek endorse government intervention similar to what Rawls's transfer branch carries out (e.g., Zwolinski 2019). Indeed, the only system of political organization the authors can think of that might not rely on expertise to an objectionable degree would be a *very* minimal libertarianism, where government does nothing more than to protect citizens against violence, theft, and fraud, and to enforce contracts (Nozick 1974: 26). Since past societies have carried out these functions without an expert bureaucracy, we believe they can be carried out in a

minimally competent way absent expert reasoning.⁹ Given the arguments of this chapter, the Nozickian night-watchmen state may be where the ardent defenders of public reason find themselves.¹⁰

§6. Objections

6.1. Scope and Delegation

A first objection to our argument rejects our position on the scope of public reason. In §2 we followed Quong (2004) and stipulated that public reason should apply any time coercive policies are at issue in the public sphere. If public reason has a narrower scope, though, then perhaps we need not worry about expert reasoning employed by administrative agencies. If public reason only applies to constitutional questions (as Rawls sometimes suggests), then agencies don't violate public reason's demands when formulating (non-constitutional) regulatory policy with their technical expertise.

Narrowly restricting the scope of public reason without a compelling justification seems objectionably *ad hoc*. Why insist on a publicly justified constitution while ignoring the public justification of specific policies? There are two broad responses. First, perhaps ordinary coercive

⁹ That said, these functions would probably be carried out in a maximally competent way aided by expert social science—for instance, by looking at the effects of different kinds of incarceration programs on recidivism. To adhere to public reason, such considerations must be eschewed and thus a less effective scheme of punishment adopted. By contrast, we cannot imagine even minimally competent environmental regulation or macroeconomic engineering absent expert reasoning.

¹⁰ Some might wonder what the difference is between our thesis and one advanced by Gaus (2010), who argues that justificatory liberalism tilts toward classical liberalism. There are many. First, Gaus does not work within the public reason framework, but rather works within a different justificatory liberalism framework. We discussed this distinction in footnote 2 above. Second, Gaus thinks that his liberal framework tilts toward classical liberalism, not libertarianism. But classical liberalism consists of much more than the night-watchmen state: "classical liberals also typically endorse the provision of public goods and improvements, education, poor relief, as well as financial, health, and safety regulations" (Gaus 2010: 235). As we noted in the paragraph above, this kind of liberalism also relies on expertise, and is thus also blocked by public reason. Finally, Gaus does not argue that his justificatory liberalism blocks egalitarianism, but rather makes this form of political organization "more, not less, difficult to justify than a more limited government authority" (Gaus 2010: 238). By contrast, we have argued that public reason does block egalitarianism, full stop.

policies are just not the kind of thing requiring public justification, whereas constitutions are. This response seems obviously incorrect. Indeed, many ordinary policies have a far greater influence on one's life than most constitutional provisions—it would be absurd to say that current firearm policies are not the sort of thing requiring justification, whereas Article V does stand in dire need of justification. A second, more interesting response says that by publicly justifying the constitution, any piece of legislation passed under that constitution is *also* publicly justified. In other words, the property *being publicly justified* carries over from the constitution to any policy passed in accordance with its procedures. Thus, if policies issued by administrative agencies are passed in accordance with a publicly justified constitution, then so too are they justified, regardless of the kind of reasoning that underlies them.

There are obvious counterexamples to this view. Constitutions that are plausibly publicly justified often do not prohibit laws that are most certainly not publicly justified, because those constitutions only set very basic rules of governance. Many people, for instance, accept the broad contours of the United States of America's Constitution, but would seriously balk at the idea of policies like a border wall being publicly justified. So it is simply incorrect to say a constitution's public justification *always* carries over to those policies passed under it. A more subtle view would go something like this: for any policy *p* passed under a publicly justified constitution, there is a *presumption* that *p* is publicly justified, but this presumption can be overturned if citizens have sufficient reason for rejecting *p*. So, all policies passed under the Constitution of the United States are presumptively publicly justified, but since objectors have strong reasons to reject a border wall, this *specific* policy is not. Why endorse such a view? One might argue for the presumption by pointing out that, by hypothesis, *p* was enacted democratically and did not violate basic constitutional rights. Citing the literature on democracy's desirable epistemic properties (e.g., Goodin and Spiekermann 2018), perhaps this gives us *pro tanto* reason to accept *p*.

Although we are skeptical about the view that policies passed under a publicly justified constitution are presumptively justified, let us grant it *arguendo*. Whether it can save the administrative state depends on what counts as a sufficient reason to defeat a policy's presumptively justified

status. There are likely numerous ways this can happen. Clearly one such way is to simply reject as bad reasons those reasons given in defense of p . Consider an example: suppose the main reason given in defense of a border wall is that most illegal immigrants are dangerous criminals. Since it was passed under a publicly justified constitution, the law is presumptively publicly justified. But, since many people reject the proposition that most illegal immigrants are dangerous criminals, it follows that the law's presumptive status is defeated.

If rejecting the reasons in defense of p is one way to strip p of its presumptive status as publicly justified, then the current proposal cannot solve controversy failure. For controversy failure occurs just when persons reject the science and social science experts offer in defense of policies. Hence, publicly justifying the constitution will not remedy at least half our problem, *even if* we grant that a publicly justified constitution presumptively justifies policies passed under it. For such presumption can be overcome if one rejects the underlying reasons in defense of a policy, and scientifically competent citizens do this with policies passed by experts at administrative agencies. Hence, these policies are not publicly justified.

Perhaps it is a partial victory, though, if restricting the scope to a publicly justified constitution solves ignorance failure. After all, those subject to ignorance failure cannot reject the reasons given in defense of the presumptively justified p , because these citizens by definition don't understand them. But rejecting the underlying reasons given in defense of p is not the *only* way of defeating p 's status as presumptively justified. Consider an example. Suppose Rowena, wanting to get in shape, hires a personal trainer. Because she freely chose to do this and because the trainer is an expert, it is plausible that everything the trainer tells Rowena to do is presumptively justified—unless she has sufficient reason to reject the trainer's advice, she ought to do what he says. Now suppose the trainer advises Rowena to do something that she reasonably suspects might cause her serious pain. When she asks why she should do this, the trainer launches into complicated kinesiology that Rowena cannot understand. By hypothesis, she cannot reject the trainer's reasons that underly his prescription. Nonetheless, it still seems that Rowena's

concerns about the advice itself defeat the trainer's presumptive claim to justification, *even though* she cannot understand his expert reasoning.

More generally, it seems that another way of defeating presumptive justification occurs when persons do not understand the reasoning in favor of the presumptively justified policy, but nonetheless have significant and reasonable concerns about it. This happens frequently. Though one might not understand the technical details behind Medicare-for-all proposals, one can have reasonably grounded concerns about such a policy given the massive scope of the intervention, the complexity of the relevant market, failures of major healthcare overhauls in the past, and so on. These sorts of cases, though, will allow citizens suffering from ignorance failure to block presumptively justified policies. If Colin has reason to reject *p*, then this suffices to strip *p* of its presumptive status, *even though* he does not understand the technical reasons given by experts in favor of *p*. This is just like the case of Rowena, who defeats the trainer's presumptive claim to justification by having reasonable concerns about his advice, *even though* she didn't understand the expert reasoning behind it.

Note that, on this view, not *all* instances of ignorance failure block administrative policies as would be the case with a conception of public reason that applies to all exercises of political power. On the latter, standard view, all instances of ignorance failure block policies, because there are no shared reasons in defense of them. But if one goes with the narrower view being considered in this section, then ignorance failure only blocks a policy when the citizen who does not understand the reasons supporting the policy *also* has reason to reject the policy. If a citizen does not understand the reasons behind presumptively justified *p*, but also has no strong objection to *p*, then *p* would retain its presumptive justified status on the narrow view. So narrowing the scope of public reason does have some positive impact, but not much.

To sum up: narrowing the scope of public reason will not do much to circumnavigate our conclusions. The best a narrow view can say is that policies passed in accordance with a publicly justified constitution are presumptively justified. This, though, will not prevent controversy failures, and will only prevent *some* ignorance failures.

A related objection to the one just considered concerns the possibility of publicly justifying *delegation*. Delegation refers to a legislative body's handing over some of its decision-making powers to administrators, who then craft policy. Legislatures often do this for an obvious reason: many policy questions require expertise, which legislatures lack. Though each and every policy an administrative agency passes may not be publicly justified (because public reason falls silent), perhaps it is sufficient to publicly justify the institution of delegation itself. If there are shared reasons justifying delegation, then we need not worry about publicly justifying those policies passed by an agency to which power has been delegated.

The structure of this objection is similar to the one just considered. With both objections, publicly justifying an institution that in some sense leads to policy p —be it a constitution or delegation—will then publicly justify p itself. Given the structural similarity, the delegation objection fails for the same reasons. If p was passed by an agency to which authority was delegated in a publicly justified manner, the best we can say is that p is presumptively justified—perhaps, as with the democratic process, because of the administrative state's epistemic properties. One way of defeating p 's presumptive status is to reject the reasons given in defense of p . Hence, controversy failure still blocks p . Another way of defeating p 's presumptive status is for persons to reject p itself, even if they do not understand the reasons given in defense of p . Hence, some instances of ignorance failure will still block p . As such, publicly justifying delegation will not do much to resolve our problems.

6.2. Idealization

Another objection goes something like this: it is not that citizens must *actually* accept or endorse the scientific and social scientific considerations that are used to justify policy p in order for p to be justified. Rather, in order for p to be justified, it must only be that citizens, were they to engage in a sufficient amount of good reasoning, would come to accept or endorse the scientific and social scientific considerations used in justifying p . That is, we *idealize*. So on this view, it does not matter

that Colin does not understand the partial equilibrium model of a market with asymmetric information sets between firms and consumers that Rowena uses in her justification of policy *p*. What matters is that, were Colin to engage in a sufficient amount of good reasoning, he would then come to accept or endorse such a consideration. If this counterfactual test is satisfied, then *p* is justified to Colin *even if* actual Colin does not accept the complex economic reasoning that Rowena employs.

There are several issues with this objection. First, we fully grant that a public reason account of justification will likely require *some* kind of idealization if *any* policy is to be justified at all. This is something detractors of public reason love pointing out (e.g., Enoch 2015). But we also firmly hold that, in order for idealization to be consistent with the underlying theoretical goals of public reason in the first place, this idealization will need to be significantly constrained. Our goal, recall, is to justify exercises of political power. Such being the case, surely idealized Colin must in some sense be accessible to actual Colin, for it is actual Colin who power is being exercised over. That is, were actual Colin to reflect on the reasons idealized Colin has, then actual Colin would come to see that he has these reasons too. Indeed, this is the standard of idealization that Gerald Gaus employs: “The reasons you have must be accessible to you, and as a real rational agent in a world in which cognitive activity has significant costs, rationality does not demand one keep on with the quest to discover less and less accessible reasons” (Gaus 2011: 253). But, as we showed in §4, the administrative state employs reasoning of the kind learned in PhD programs or other professional schools. It is hard to see how such considerations would be accessible to the average citizen in a world in which cognitive activity has significant costs, especially since understanding *all* the reasons used by *all* agencies would require more PhDs than one could acquire in a lifetime.¹¹

¹¹ Perhaps one could reject our use of Gaussian-style idealization. But then what standard of idealization would one employ? On the standard Rawlsian line, idealization attributes substantive moral values and commitments to persons (they see themselves as free and equal, desire to participate in a fair system of social cooperation, etc.). It is possible that idealization could *also* attribute scientific and social scientific views to people as well. But in order for this strategy to work, idealization must attribute to persons several PhDs worth of knowledge. This seems more controversial than, say, attributing to persons substantive Catholic or Protestant views, which is at odds with the liberal project more generally.

Second, the proposed idealization-based solution only purports to rescue the administrative state from ignorance failure, not from controversy failure. Recall our distinction from §3. With ignorance failure, the problem is that most citizens do not understand the kinds of reasoning employed by administrative agencies. It thus cannot be said that they accept or endorse such considerations. With controversy failure, however, the problem is that citizens who *do* understand the kinds of reasoning employed by the administrative state could quite reasonably disagree with such reasoning. As Vermeule has noted, agencies often make policy in areas where there is no scientific or social scientific consensus. Because this is the case, a “general” and “well-settled” administrative law principle is that “when experts disagree, agencies are entitled to rely upon the reasonable opinions of their own qualified internal experts” (Vermeule 2016: 132–133). But of course, other experts can quite reasonably reject such opinions. Thus, *even if* idealization could solve ignorance failure (we doubt that it does), it could not solve controversy failure.

6.3. Learning and Deliberation

Another response to our central argument is that, though normal persons with normal cognitive capacities cannot understand the complicated reasoning employed by administrative agencies, perhaps citizens can come to learn these reasons. This could be through engaging with public intellectuals who are experts in the relevant fields (as an example, citizens could read op-eds penned by these intellectuals); participating in the public comment processes on proposed rules; or even spending designated civic holidays deliberating about proposals (Ackerman and Fishkin 2004). According to the proposal under consideration, these processes will ensure that citizens learn the relevant scientific and social scientific reasons. The demands of public reason are thus satisfied.

As a first response: while all these processes will surely improve citizens’ knowledge in areas outside their expertise, we doubt they are sufficient to solve ignorance failure. This is because the moderate amounts of education they involve are no replacement for the years of

study drawn on by experts. For example, we don't doubt that lay readers can improve their knowledge of economic policy by reading Paul Krugman's *New York Times* columns, but that does not make them competent to understand, say, the reasons for or against a certain trade policy that Krugman would employ were he to advise the Department of Commerce. Indeed, if it did, several aspects of the administrative state we discussed in §4 would be inexplicable. This point is not controversial. Moderate amounts of reading and deliberation are not equivalent to sustained education and professional development.

But suppose for the sake of argument that there were sufficient formal and informal mechanisms for all citizens to learn all scientific and social scientific reasons supporting all administrative policies. This, like idealization, would not solve controversy failure. But not only would this purported solution to ignorance failure not *solve* controversy failure, it would likely *exacerbate* it. Citizens who come to learn the relevant scientific and social scientific reasons will likely come to disagree about those reasons, just as experts do. For example, most people who do not know any basic economics likely do not have strong opinions on, say, the effects of regulation on financial markets. But of those citizens who come to understand basic economics through engaging with public intellectuals, some will agree with Paul Krugman on the effects of regulation, and others will side with Milton Friedman. The proposed solution, therefore, will cause many citizens who initially did not understand the justifications for coercive policies to come to reject them, and these policies would therefore remain unjustified. There is no reason to believe that a public reason proponent would view shifting ignorance failures to controversy failures as an improvement.

6.4. Means versus Ends

A final objection goes something like this: the administrative state is only concerned with the implementation of policy, not the actual crafting of policy. That is, the administrative state is only concerned with the *means* of implementing policy, not the actual *ends* of policy itself. And, one might argue, in order to justify policy *p*, public reason must only be used

in justifying the ends of p , not the means of achieving p . So when it comes to justifying the ends of policy—that we want a single-payer healthcare system, or a property-owning democracy, or a basic income—we must proceed from shared reasons. But when it comes to the nitty-gritty of implementation—figuring out the demand elasticity of health insurance, for instance—public reason is inapplicable, and we do not need to proceed from shared reasons. Since the administrative state is all about implementation and not the actual crafting of policy, it follows that the administrative state is not subject to the demands of public reason. It does not matter, then, that agencies employ reasoning subject to the ignorance and controversy failures.

Our first response to this objection is that, at least in modern constitutional democracies like the United States, although legislatures can and do constrain agencies to a certain extent, administrative agencies nonetheless play a major role in policymaking that goes beyond mere implementation of democratically decided ends. Agencies get their powers from statutes, provided those statutes are consistent with the Constitution. Thus, there are two possible limits to their authority: statutory and constitutional. In the United States, the constitutional limit is known as the “non-delegation doctrine,” which says that Congress cannot delegate its “legislative power” to an administrative agency. Since 1928, the Supreme Court has held that, as long as Congress lays down an “intelligible principle” to which the agency must conform, it does not unconstitutionally delegate legislative power (*J. W. Hampton, Jr.* 1928: 409).

But this standard, as it has been developed, is consistent with agencies wielding what most would consider to be large amounts of policymaking authority that, since agencies’ power is given by statute, also gives a sense of the limited extent to which statutes typically limit agencies’ power. For example, in *Whitman v. American Trucking Associations, Inc.*, the Supreme Court upheld, against a non-delegation challenge, a provision of the Clean Air Act that instructed the Environmental Protection Agency (EPA) to set “ambient air quality standards the attainment and maintenance of which in the judgment of the Administrator, based on [the] criteria [documents of §108] and allowing an adequate margin of safety, are requisite to protect the public health” (*Whitman* 2001: 472).

These “criteria documents” were to be created by the EPA based on certain open-ended considerations established by Congress, which included “any known or anticipated adverse effects on welfare” (42 U.S.C. § 7408(a)(2)(C)). In essence, then, the statute granted the EPA the power to set whatever air quality standards were necessary to protect public health, and gave the EPA essentially full leeway to decide when public health was protected¹²—based, of course, on “the latest scientific knowledge useful in indicating the kind and extent of all identifiable effects on public health or welfare which may be expected from the presence of such pollutant in the ambient air, in varying quantities” (42 U.S.C. § 7408(a)(2)). If this isn’t policymaking about ends, we don’t know what is. But if the reader is not convinced by even this, note that *Whitman* observed that, in prior cases, the Supreme Court had found an intelligible principle “in various statutes authorizing regulation in the ‘public interest’” (*Whitman* 2001: 474). And the Court itself noted that constitutional grants of power to agencies give them a “permissible degree of policy judgment” (*Whitman* 2001: 474). Clearly, then, administrative agencies focus on both the means *and* ends of policy.¹³

Of course, we don’t mean to suggest that the current state of constitutional law in the United States is conclusive of agencies’ proper role in a democracy.¹⁴ Perhaps the American system is simply impermissible.

¹² Of course, when a statute limits the reasons an agency can consider, the agency must follow the statute, and one holding of *Whitman* was that the statute prohibited the EPA from considering the economic costs of regulating air pollution. But here it is worth noting that American courts give agencies substantial deference in interpreting statutory terms under the so-called *Chevron* doctrine, which, in practice, gives them even more policymaking power because it allows them to declare, within limits, what the law means, and hence which factors they may consider. See Siegel (2018: 960).

¹³ Richardson also rejects a strict ends-means division between legislative and agency functions, and agrees that agencies do and should specify certain ends, but argues that certain forms of public participation in agency rulemaking, combined with legislative oversight, public input into the appointment of agency heads, and the enhancement of agency professionalism, allow agency deliberation to “count as a continuation of fair processes of democratic deliberation” (Richardson 2002: 214–230). While this might suffice to solve the problem Richardson seeks to tackle—that of bureaucratic domination—we do not take Richardson to be arguing that this suffices to publicly justify the agencies’ policies in the sense public reason theorists like Rawls are concerned with. Nor would his solutions so suffice. Public participation might slightly mollify ignorance failure, but would not solve it; indeed, it might even exacerbate controversy failure. And none of Richardson’s other solutions address the problems we have been concerned with.

¹⁴ Although conventional wisdom says that the non-delegation doctrine was effectively repudiated by the New Deal, recent research supports the conclusion that “there was never a

Suppose then that the administrative state was not actually involved in crafting policy; it only worked on the implementation side of things. Our second response is simply that, intuitively speaking, justifying a broad, vague “end” of policy is not sufficient to justify the actual, implemented, coercive policy. For example, suppose—borrowing from the case of *Burwell v. Hobby Lobby Stores, Inc.*—that Rowena is a devout Catholic and business owner. Colin argues—strictly using reasons that Rowena accepts—that she ought to accept policy *p*, whose end is affordable healthcare for all. Any coercion stemming from *p* would thus, on the current objection, be justified to Rowena. Yet, when the details of implementation are fleshed out by HHS, Rowena’s business is compelled by law to provide contraceptives to female employees, which is in violation of her faith and conscience. We do not think that such a coercive interference is publicly justified to Rowena (given her religious commitments), *even though* by hypothesis the vaguely specified end that the regulation serves was justified to her with shared reasons. Thus, *even if* it were true that administrative agencies were strictly about implementation of policy ends (they are not), it still does not seem sufficient to justify only the ends of policies to persons in order to justify all coercive aspects of the policy itself. The devil is often in the details, and these details also stand in need of public justification.

§7. A Concluding Section

Public reason theorists argue that political power must be justified over those whom it is exercised. This requires that those subject to political power accept or endorse the reasons given for its justification. These reasons include scientific and social scientific considerations. In the modern administrative state, agencies do and must use complex scientific considerations to make policy. This results in ignorance failure and controversy failure. These failures prevent the policies from being justified to most citizens. As a result, public reason condemns the administrative

time in which the courts used the nondelegation doctrine to limit legislative delegations of power” (Whittington and Iuliano 2017: 381).

state and political theories that presuppose it, such as egalitarianism, as unjustified exercises of coercive force.

To the extent that theorists advocate public reason as a means of reconciling state power with citizens' freedom, our argument raises a new way that this conception of freedom conflicts with equality. This conception of freedom conflicts with equality because it condemns the only (plausible) way of achieving it. Indeed, it could very well turn out that public reason permits little more than the Nozickian minimal state and the massive distributive inequality that might ensue. If so, properly understanding the demands of public reason could very well change existing fault lines in the current literature. Egalitarians might be driven to reject public reason, and libertarians to accept it.

Works Cited

- Ackerman, Bruce and James S. Fishkin. 2004. *Deliberation Day*. New Haven, CT: Yale University Press.
- Arrow, Kenneth J. and Gérard Debreu. 1954. "The Existence of an Equilibrium for a Competitive Economy." *Econometrica* 23: 265–290.
- Badano, Gabriele and Matteo Bonotti. 2020. "Rescuing Public Reason Liberalism's Accessibility Requirement." *Law and Philosophy* 39: 35–65.
- Banerjee, Abhijit, Esther Duflo, and Rachel Glennerster. 2010. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6: 487–500.
- Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 7: 22–53.
- Boettke, Peter J. 1997. "Where Did Economics Go Wrong? Modern Economics as a Flight from Reality." *Critical Review* 11: 11–64.
- Citizens to Protect Overton Park v. Volpe*, 401 U.S. 402 (1971).
- Daniels, Norman. 1981. "Health-Care Needs and Distributive Justice," *Philosophy & Public Affairs* 10: 176.
- Deaton, Angus and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Control Trials." *Social Science & Medicine* 210: 2–21.

- Enoch, David. 2015. "Against Public Reason." In *Oxford Studies in Political Philosophy*, Vol. 1, edited by David Sobel, Peter Vallentyne, and Steven Wall. Oxford: Oxford University Press.
- Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.
- Gaus, Gerald. 2010. "Coercion, Ownership, and the Redistributive State: Justificatory Liberalism's Classical Tilt." *Social Philosophy & Policy* 27: 233–275.
- Gaus, Gerald. 2011. *The Order of Public Reason*. Cambridge: Cambridge University Press.
- Gifford, Daniel J. 1984. "The New Deal Regulatory Model: A History of Criticisms and Refinements." *Minnesota Law Review* 68: 299–332.
- Goodin, Robert E. and Kai Spiekermann. 2018. *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.
- Jønch-Clausen, Karin and Klemens Kappel. 2016. "Scientific Facts and Methods in Public Reason." *Res Publica* 22: 117–133.
- J. W. Hampton, Jr. & Co. v. United States, 276 U.S. 394 (1928).
- Kuhn, Thomas. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*: 320–339. Chicago, IL: University of Chicago Press.
- Kuhn, Thomas. 2012. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Landis, James M. 1938. *The Administrative Process*. New Haven, CT: Yale University Press.
- Larmore, Charles. 1996. *The Morals of Modernity*. Cambridge: Cambridge University Press.
- Lister, Andrew. 2013. *Public Reason and Political Community*. London: Bloomsbury.
- Meazell, Emily Hammond. 2011. "Super Deference, the Science Obsession, and Judicial Review as Translation of Agency Science." *Michigan Law Review* 109: 733–784.
- Motor Vehicle Manufacturers Association v. State Farm Mutual Auto Insurance Co.*, 463 U.S. 29 (1983).
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Quong, Jonathan. 2004. "The Scope of Public Reason." *Political Studies* 52: 233–250.

- Quong, Jonathan. 2014. "On the Idea of Public Reason." In *A Companion to Rawls*, edited by Jon Mandle and David A. Reidy: 265–280. Oxford: Wiley Blackwell.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rawls, John. 1999. "Justice as Fairness: Political Not Metaphysical." In *Collected Papers*: 388–414. Cambridge, MA: Harvard University Press.
- Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.
- Rawls, John. 2005. *Political Liberalism*. New York: Columbia University Press.
- Richardson, Henry. 2002. *Democratic Autonomy*. Oxford: Oxford University Press.
- Salanie, Bernard. 2000. *Microeconomics of Market Failures*. Cambridge, MA: The MIT Press.
- Seidenfeld, Mark. 1992. "A Civic Republican Justification for the Bureaucratic State." *Harvard Law Review* 105: 1511–1576.
- Siegel, Jonathan. 2018. "The Constitutional Case for *Chevron* Deference." *Vanderbilt Law Review* 71: 960.
- Tuomala, Matti. 2016. *Optimal Redistributive Taxation*. Oxford: Oxford University Press.
- Vallier, Kevin. 2019. *Must Politics Be War?* Oxford: Oxford University Press.
- Vermeule, Adrian. 2016. *Law's Abnegation*. Cambridge, MA: Harvard University Press.
- Wald, Patricia. 1996. "Judicial Review in Midpassage: The Uneasy Relationship Between Courts and Agencies Plays On." *Tulsa Law Review* 21: 221–258.
- Whitman v. American Trucking Associations, Inc.*, 531 U.S. 457 (2001).
- Whittington, Keith and Jason Iuliano. 2017. "The Myth of the Nondelegation Doctrine." *University of Pennsylvania Law Review* 165: 381.
- Zwolinski, Matt. 2019. "A Hayekian Case for Free Markets and a Basic Income." In *The Future of Work, Technology, and Basic Income*, edited by Michael Cholbi and Michael Weber. New York: Routledge.

An Autonomy-Based Argument for Democracy

James Lindley Wilson

Democracy is often associated with freedom. The connection is old: Plato has Socrates describe the democratic city (not without concern) as “full of freedom and free speech” (Plato 1991: 557b, p. 235); Aristotle agrees that democratic regimes are founded on a “presupposition” of freedom (Aristotle 1984: 1317a40, p. 183). Many of our contemporaries continue to believe the connection to be as deep as the Greek philosophers did. Today, whatever freedom is associated with democracy garners more unequivocal praise than the ambivalent ancients would allow. When we value freedom, then, the connection between democracy and freedom seems to promise an argument for democracy: we should establish, maintain, and perhaps obey democratic regimes, because this establishes, maintains, or honors freedom in some important way.

I will defend a variant of this idea. There is an argument for egalitarian political systems grounded in a kind of freedom, or, as I shall call it, autonomy. The argument is not a matter of accepting the Platonic-Aristotelian claim that democracy grants “license in it to do whatever one wants,” and endorsing that license (Plato 1991: 557b, p. 235; Aristotle 1984: 1317b12–13, pp. 183–4). Instead, the value of autonomy is more directly at stake in the distribution of political rule.¹ Respecting the autonomy of others requires granting them some authority over patterns of social interaction that implicate and direct their wills in ways I will specify. Directing and implicating of the wills of others

¹ This is closer to Aristotle’s observation (1984: 1317b1–2, p. 183) that democracy involves the right to rule and be ruled in turn, which is itself a kind of freedom.

without granting them any authority over the actions that so significantly shape and constrain their choices objectionably fails to respect their autonomy, understood as a claim to direct their own lives in certain central ways.² Because we ought to respect equally the autonomy of all persons, we ought to grant equal authority to each person who participates in sufficiently dense patterns of social interaction. Equal authority over the terms of social life just is political equality. So respect for the autonomy of others requires us to organize ourselves democratically, in the sense of making collective decisions in ways consistent with political equality.

An important upshot of this argument is that a democratic citizen's autonomy is in one sense respected even when the regime enacts laws and policies that contradict her judgment about what the regime ought to do. This is because, as I shall argue, each citizen's autonomy claim is a claim to equal authority over what the citizens all do together, not a claim for each citizen to control or actually determine what all do together. The argument thereby aims to solve a democratic puzzle made famous by Rousseau (1997b: Book IV, chs. 1–2): how democratic citizens can be free despite their lack of consent to particular laws.

I defend these claims by identifying the limits of other attempts to defend democracy on the basis of its connection to freedom or autonomy. I then distinguish several types of autonomy, and argue for an autonomy-based claim to authority over matters of common concern. Familiar requirements of non-interference in certain actions, and an adequate set of options to choose from, are grounded in concerns for authorship of one's life that also support claims to (shared) authority over how one's options are determined. This argument reveals gaps in otherwise illuminating discussions of freedom and its relationship to democratic or republican government—gaps typically involving the way that aggregated, unintended consequences of others' decisions can infringe on one's autonomy without constituting any bilateral relationships of domination or

² I describe this argument as “autonomy-based” rather than “freedom-based” for two reasons: first, to emphasize the grounding of these claims in entitlements to self-direction, an ideal traditionally linked with autonomy; and, second, to clarify that I do not provide a comprehensive account of individual freedom or free choice.

subjection.³ Because such consequences can direct and implicate others' wills, those others have a claim to (shared) authority over actions that produce such consequences, even if the lack of such authority would not constitute domination or subjection. I conclude by addressing the question of how substantial an action's effect on the wills of others must be in order to trigger claims to shared authority. I specify some circumstances that are sufficient to trigger shared authority claims, and which are clearly satisfied by contemporary societies. We thus have autonomy-based reasons for those societies to organize on terms of political equality.

I. Democracy and Autonomy: Alternative Approaches

We can separate autonomy-based arguments for democracy into two categories. Arguments in the first category claim that democracy is required to promote or realize the autonomy of some collective agent, such as a people or nation. Arguments in the second category claim that democracy is required to promote or realize the autonomy of individuals. I will consider each in turn.

A. Collective Autonomy Arguments

Collective autonomy arguments have the virtue of simplicity. While there are metaphysical details to work out in specifying what exactly a collective agent such as a people or state or nation is, and how it may act, philosophers have largely overcome these problems (Hobbes 1994: ch. 16; Gilbert 1996; List and Pettit 2011). Once we understand what a collective agent is and how it can act, it should not be too difficult to articulate what it means for such an agent to act autonomously—roughly, to determine itself how it acts. Democratic regimes might satisfy such criteria for autonomous action: democratic institutions might constitute a collective agent, or give expression to some otherwise-constituted

³ To preserve some readability, I keep detailed distinctions between my arguments and other prominent views of freedom and democracy in the footnotes.

agent, and democratic decision-making procedures may constitute a way for that agent to rule itself. The agent (say, the state) may determine what constitutes its beliefs and desires through democratic procedures, and executing democratic decisions may thus amount to acting according to the beliefs and desires of the acting agent in a suitably autonomous manner.⁴

Collective autonomy arguments fail to support democracy, however, for two reasons. First, there is no reason why democratic procedures are necessary to constitute autonomous collectives. Second, it is not clear why we should value the autonomy (in this sense) of collectives. Let me explain.

Collective autonomy arguments do not support democratic procedures or institutions. There is nothing in leading analyses of group agents that requires such agents to have an internally egalitarian structure in order to count as a collective agent. Nor, once a collective agent exists, is there any reason to think that it can only act autonomously if its beliefs or desires or whatever are generated in egalitarian fashion. Autonomy requires that the agent's psychological features suitably determine its actions. But for any psychological feature of a collective agent, there are non-democratic ways to generate that feature. (An autocrat may determine the state's beliefs or desires or dispositions or whatever.) And there are non-democratic ways to ensure that the relevant features determine the agent's actions. (Reliable, but non-democratic, constitutional systems may ensure that the autocrat's will determines the state's actions.) Collective autonomy, while coherently connected to collective decision procedures, does not require, or even favor, democracy.

The second problem with collective autonomy arguments for democracy is that they require some explanation of what is valuable about the self-rule of some collective agent. Defenders of individual autonomy must similarly explain why individual autonomy warrants respect of a kind that constrains how we treat one another. There is, however, a long and vibrant tradition of thought attempting to provide that explanation. I think it is fair to say that skeptical questions are much more threatening for collective autonomy arguments.

⁴ This rough sketch could be replaced by another emphasizing a different analysis of the moral psychology of autonomous agents (e.g., one without emphasis on beliefs and desires).

The reason this skepticism about collective autonomy is threatening to such arguments is that we have no reason to care about collective agents *per se*. As long as these agents possess no centralized consciousness or capacity to feel pleasure or pain or the exercise of their own agency, it is hard to see what claims the agents could have, apart from claims derived from the claims of individual agents who bear some relationship to the collective agent. A state may have claims to certain treatment on the grounds that such treatment respects the rights and interests of its citizens, for example, but this is precisely to derive the collective agent's claims from the claims of individuals. I am suggesting that all of the claims of collective agents—including any claims to autonomy—are derivative in this sense: a collective agent only has claims to certain treatment if those claims are appropriately connected to the claims of individuals.

I am not really arguing for this claim. Perhaps someone could develop a convincing argument for the intrinsic significance of collective autonomy. More likely, collective autonomy might be valuable because it serves some other values grounded in the claims of individuals. Arguments for national self-determination can be of this kind (see Stilz 2019). Such arguments would provide further justification for the claim that we ought to develop and respect collective autonomy. If these arguments succeed, however, it would not be collective autonomy that would be supporting claims to democratic political organization. The further justification would be providing the real support. This further justification would either make reference to individual autonomy, or it would make no reference to autonomy at all. So the idea of collective autonomy would not play a direct role in justifying any particular political regime.

B. Individual Autonomy Arguments

Individual autonomy arguments have the advantage over their collective counterparts that they can make use of well-developed defenses of the importance of individual autonomy. The challenge for such arguments is to make a convincing case that such autonomy has a deep and important connection to democratic political organization. Despite the long

pedigree of arguments connecting democracy and freedom, this challenge is not easy to overcome.

The simplest version of an individual autonomy argument for democracy holds that democratic societies tend to respect individual autonomy better than non-democratic societies, because the outcomes of democratic procedures involve such respect more reliably than the outcomes of non-democratic procedures. Democratic societies might best respect the legal and moral rights to which autonomous individuals are entitled, for instance, because democratic laws tend to establish and support such rights, and because democratic procedures educate or habituate citizens into respecting such rights. There is some evidence to support this claim (Christiano 2011). So there may be good, pragmatic reason for those who value individual autonomy to prefer democracies.

Pragmatic, instrumental reasons can be very good reasons. But this simple argument for democracy has limits. It is possible for democratic collectives to restrict individual autonomy through democratic means—for example, by passing, through pristinely egalitarian procedures, laws that restrict basic liberties. Democracy and autonomy can, in particular cases, be at odds.

Neither the possibility nor the reality of democratic autonomy violations defeats the simple argument. Democracy may be preferable to alternatives on autonomy grounds despite quite serious imperfections. But the real risk of democratic autonomy violations limits the significance and scope of the simple argument. First, it renders arguments for democracy vulnerable to challenges from less democratic alternatives that might do better than existing democracies at protecting and promoting individual autonomy. Even if existing non-democratic states are worse than democracies on this score, there could be “liberal-aristocratic” regimes—politically unequal regimes committed to individual rights protection—that better protect autonomy than either existing democracies or existing non-democracies (Arneson 2004). The simple individual autonomy argument for democracy is weak in that it fails to reject convincingly all non-democratic alternatives.⁵

⁵ Even if one is confident that no regime repudiating democracy (such as rule by a small aristocracy selected by exam) could better protect autonomy than a democracy, one should still

The simple argument is also weak in a second way. The simple argument supports democracy only so long as democratic procedures actually produce outcomes that involve better respect for autonomy than do the expected outcomes of non-democratic procedures. Contingent, purely instrumental defenses of democracy suffer the defect that their general acceptance by citizens undermines some of the values that democracy otherwise achieves, by rendering citizens' commitments to equal relations less robust. Citizens who are willing to abandon democracy for better outcomes are less committed to the equality that democracy involves, and their relations therefore partake less in whatever goods that equality constitutes or brings.⁶

This second weakness is not dispositive: other considerations may weigh so heavily in favor of the simple argument that we should accept the argument despite some bad consequences of citizens coming to publicly rely only on this argument. But the weakness suggests that there would be value in a firmer connection between democracy and autonomy.

We should focus that search on the respect for autonomy (if any) embodied in democratic procedures themselves. One argument states that democratic procedures respect autonomy in that they allow all individual citizens to exert control over what the citizens, collectively, do. When citizens exert democratic control over laws and policies, the argument goes, each citizen's psychological states (desires, beliefs, wills, whatever) exert some control over what all citizens, including the citizen herself, will do. This, the argument continues, is a valuable kind of autonomy, a rule over self (and shared rule over others). This argument follows Rousseau's thought that democracy enables moral freedom: the state of being subject only to laws one gives oneself (Rousseau 1997b: Book I, ch. 8, Book II, ch. 4).

worry about the simple argument's inability to sustain criticisms of smaller-scale political inequalities—such as small inequalities in voting weights, inequalities in deliberation and informal influence, or elite-dominated constitutional restrictions on democratic action—within a wider, vaguely democratic (e.g., mass-electoral) framework. If the simple argument provides little support for such criticisms of political inequality, it is a limited argument for democracy, if the latter is understood as a regime reflecting full political equality.

⁶ For development of this argument, see Wilson unpublished a. On democracy and relations of equality, see Wilson 2019: chs. 1–2; Kolodny 2014b.

I will defend a version of this argument. But there are defective versions of the argument we should reject. Thomas Christiano (1996: ch. 1) argues that defenses of democracy grounded in “self-government” fail. (See also Viehoff 2014: 350–1.) Christiano’s argument, roughly summarized, has three prongs: first, self-government through democratic participation tends to come at the expense of self-government in private pursuits, in that everyone is subject to government regulation of their private lives; second, this self-government through democracy is less valuable than private self-government, since the latter tends to contribute much more directly and broadly to the overall shape of one’s life; and, third, disagreement between citizens makes the democratic self-government of citizens incompatible—some will successfully self-govern, but others will not, as they will be outvoted.

Christiano is right that democracy does not guarantee to any individual self-government in the sense of the ability actually to determine through democratic procedures what people (including that individual herself) will do with respect to a wide range of issues determining the shape of her life. Even when there is concordance between one’s judgment and the collective action, there is only a limited sense in which that concordance is because of one’s judgment (though cf. Tuck 2008). Government by many people, even on equal terms, cannot guarantee that level of individual control over what people do—over the outcomes of collective decisions.⁷ Understanding self-government or autonomy this way, then, Christiano is correct that democracy cannot deliver self-government.⁸

There is another important sense, however, in which democratic procedures do respect the autonomy of citizens. Respect for autonomy does not require granting to each the kind of substantial control that

⁷ For a similar conception of control as determining collective outcomes, see List and Pettit 2011: 136.

⁸ This poses a problem for Philip Pettit’s (2012: ch. 3) freedom-based defense of democracy. For Pettit, state interference with a citizen’s choice would not count as dominating if the citizen controlled that interference. Democracy, says Pettit, grants individual citizens equal, fractional shares of collectively efficacious control over state interference. But because this fractional share is insufficient on its own to control the state, it is not clear why it counts as the kind of control over interference necessary, on Pettit’s account, to preclude domination (and thus preserve freedom).

Christiano imagines. We can respect autonomy by sharing authority, rather than attempting, quixotically, to universalize absolute self-rule. I begin the defense of this claim by clarifying what it is to make autonomy-based demands of others.

II. Aspects of Autonomy

Nomy Arpaly (2003: 117, 118) describes autonomy as an “overworked term.” Often when philosophers discuss autonomy, they mean some kind of internal self-rule, according to which one’s actions are appropriately determined by some of one’s relevant psychological states. So, for instance, one might be autonomous if one’s reason, rather than one’s desires, determines one’s action; or one might be autonomous if one’s higher-order desires (desires about what to desire) align properly with one’s lower-order desires (desires whose objects are not desires). Sometimes, however, as Arpaly (2003: 120) points out, by autonomy we mean “normative” or “moral” autonomy, namely moral claims to make one’s own decisions without interference or control by others. (Normative autonomy may also involve claims to assistance, for instance to aid in developing capacities necessary to make one’s own decisions.) Whereas other conceptions of autonomy are matters of the internal constitution of one agent, normative autonomy is interpersonal: normative autonomy grounds claims to certain kinds of treatment by others. As Arpaly (2003: 120–1) notes, one may have legitimate claims to normative autonomy even if one is not autonomous in one of the several internal senses of autonomy. For instance, a doctor typically owes a patient respect for his (normative) autonomy in dispensing medical treatment even if the doctor knows that the patient’s decision will not be autonomous in the sense that it will not reflect the patient’s reason or better judgment or higher-order desires or whatever. Similarly, one may violate another’s normative autonomy without undermining her autonomy in the internal sense: even if I steal from you, or assault you, or imprison you, you may well remain autonomous in the sense that you could respond to any of those wrongs in a way that accords with your reason or true self or whatever (Buss 2005). Normative autonomy—what we

owe to persons as a matter of respect for their autonomy—is not a matter of ensuring internal autonomy in the sense of a proper connection between an agent’s actions and certain of her psychological features. Respecting normative autonomy is not sufficient, and not always necessary, to ensure an agent’s internal autonomy; nor is an agent’s internal autonomy necessary (though it may be sufficient) to entitle her to the claims of normative autonomy.⁹ This is true even if the justification for the claims of normative autonomy has something to do with the claiming agent’s capacity for internal autonomy.

In developing an argument for democracy, we are concerned with normative autonomy—what kinds of claims autonomous or potentially autonomous agents have on us. Normative autonomy centers on an agent’s claims, with respect to others, to determine her actions herself, as opposed to having the actions determined by others, or by nobody at all (as perhaps in cases of extreme, naturally caused duress). The ideal of self-direction is what connects normative autonomy with internal autonomy, though the latter will typically involve more stringent requirements on what counts as truly *self-direction* (Darwall 2006). Our question, then, is what claims normative autonomy supports.

III. The Democratic Demands of Normative Autonomy

In this part, I argue that respecting persons’ autonomy requires granting them authority over common matters that shape their exercise of choice. They have a claim to authority on equal terms with others whose choice is also shaped by these matters. I begin with relatively uncontroversial claims about the requirements of normative autonomy, and argue that the reasons supporting these claims support my thesis that respecting autonomy requires equally sharing authority over common life—that is, regulating common life democratically. These reasons are grounded in individuals’ claim to self-direction: to determine their choices and so to

⁹ Some normative autonomy may be necessary for internal autonomy, because some forms of interference do make internal autonomy impossible, and because some forms of assistance (most notably, child-rearing) are necessary for an agent to develop the capacity for autonomy.

shape the course of their lives, as much as this is consistent with the similar claims of others. As we shall see, others' actions often threaten these claims to self-direction even when those actions do not directly interfere in one's choice.¹⁰ So respecting autonomy claims to self-direction requires granting individuals some authority over the actions of others that otherwise would, even absent strict interference, unduly direct them.

A. From Non-Interference to Adequate Spheres of Concern

Typically, when we think of normative autonomy and its claims, we imagine cases in which individuals are entitled to engage in some actions without direct interference by others. Interference is direct when it involves coercion, manipulation, or the actual use of force (if one does not consider this last to be a type of coercion).¹¹ Exactly how to define coercion or manipulation is challenging, but the details are not relevant here. Call one's "sphere of concern" the set of actions an agent has a right to undertake without direct interference by others. To endorse normative autonomy is to assert that our spheres of concern are not empty. Relatively uncontroversial examples of what belongs in everyone's sphere of concern include determining, within wide limits, what one does with one's body, for example when it comes to invasive medical care, deciding what to eat or drink, or enjoying freedom from assault. Or one might consider one's rights to determine how to dispose of one's property; for example, determining what happens to one's apple or determining who enters one's bedroom.¹²

¹⁰ As will become clear, this is distinct from the republican concern about capacity for interference. Others' actions can threaten our self-direction without constituting either interference or domination.

¹¹ Pettit (2012: ch. 1) has a broader conception of interference than this loosely Kantian notion, and, of course, denies that interference in his sense necessarily violates freedom. Even for Pettit, however, interference in choice is an atomic part of unfreedom, since domination requires the capacity to interfere.

¹² For further discussion of the idea of one's sphere of legitimate authority, see Shiffrin 2000; Ripstein 2006.

One may exercise internal autonomy, and perhaps a morally salient kind of freedom, when one chooses how to react to being tied to a chair in a prison cell. But normative autonomy requires more than this minimal freedom: it requires protection of a wider sphere of concern within which one can act, such as a reasonably wide scope for free movement. That is, respecting autonomy requires respecting others' minimally adequate spheres of concern (Raz 1986: 373–7; Pettit 2012: ch. 2).

This sufficiency requirement involves more than identifying categories of actions that are protected against direct interference. It is not enough, for instance, to define one's sphere of concern solely with reference to (independently determined) rights against physical assault and rights to property.¹³ Consider, for instance, two neighbors whose land entirely surrounds the (modestly sized) property of a third neighbor. If the two neighbors build impenetrable walls around the third neighbor's house and gardens while she sleeps, I judge that they violate that neighbor's autonomy. They do this by excessively reducing her sphere of concern, despite (by hypothesis) not directly interfering with any property rights or rights against assault. The example illustrates, I believe, that autonomy concerns establish claims to an adequate sphere of concern. Autonomy does not merely require independent claims against certain types of interference, which together constitute spheres of concern that may, like our third neighbor's, end up with any manner of extent and quality due to the actions of others.¹⁴ This means that normative autonomy will often require more than respecting negative claims against direct

¹³ "Independently determined" here means determined without reference to the requirements of sufficient autonomy under consideration.

¹⁴ I am not sure if Arthur Ripstein should be interpreted as disagreeing. He argues that external freedom requires public roads, to prevent "land-locked property" cases. Such public elements are necessary in order to prevent citizens' voluntary interactions from being subject to the decisions of other property owners (Ripstein 2009: 245–6). But one's ability to voluntarily interact with others is regularly subject to the decisions of others. This happens when we do not have adequate resources to interact, and others use their property rights to refuse to fund our interaction. If Ripstein is effectively arguing that one way of being wrongly subject to the decisions of others occurs when one has an inadequate sphere of concern, I agree. If the concern about subjection makes no reference to adequacy, then I am not sure it can support his conclusion that non-subjection does require public roads but does not require egalitarian resource distribution. The public roads case thereby calls into question how to draw the distinction between wronging a person and merely changing the context in which that person acts (Ripstein 2009: 241). The lack of public roads provision wrongs people by affecting them in

interference, because the result of many people's actions may, consistent with respecting those claims, result in undue diminution of one's sphere of concern. They may, that is, effectively build walls too narrowly around others.

B. Determining Spheres of Concern

I believe the demands of normative autonomy do not stop with this requirement of adequate spheres of concern. Respecting a person's autonomy also requires respecting her claims to authority over how her (adequate) sphere of concern is determined by others. Typically, many people will have such authority claims with respect to the same matters, because many people's spheres of concern are determined in part by those matters. Furthermore, if agents are equally entitled to respect for their autonomy (e.g., because they are moral equals in some fundamental sense), respect for autonomy is owed equally to each implicated agent. Democracy constitutively involves recognizing, equally and simultaneously, the authority claims of each individual implicated by some common matter, and thus recognizing this aspect of normative autonomy (Wilson 2019: chs. 2, 4; Kolodny 2014b: 295). So moral equality, and the claim to normative autonomy over matters of common concern, together require political equality—that is, democracy. There is an autonomy-based argument for democracy if we can establish that there are common matters that demand shared authority.

Before proceeding, a word about what it means for a person to have authority over terms of interaction. By “authority” here I do not mean the power to obligate others to obey one's commands. (It is not clear how all citizens could equally have this power.) Instead, I mean authority in a broader sense of entitlement to what I call *consideration*: recognition by others that one's decisions about what others ought to do create obligations for those others to deliberate or act in ways that take positive account of those decisions. At the limit, this recognition could amount

ways that also sometimes arise when others change the context in which one acts. I submit that a criterion of adequacy of spheres of concern is necessary to preserve the desired distinction.

to obedience to the commands of another—one person’s right to set unilaterally the terms of some interaction, say. But this broad conception of authority allows for sharing: for mutual extensions of consideration among many people with respect to some collective choice. Democracy involves granting such consideration to each citizen on equal terms. Exactly what procedures secure such equal consideration is, of course, an important question; I treat it at length elsewhere (Wilson 2019: chs. 5–6).

Why do people have autonomy-based claims to authority over how their spheres of concern are determined? The intuitive idea is that the same interest in or entitlement to self-direction that underpins what I will call “first-order” autonomy claims to determine actions without direct interference within one’s sphere of concern also supports “second-order” autonomy claims to shape the content of that sphere.¹⁵ If not a logical inconsistency, there is an untenable tension between demanding respect on grounds of self-direction in the first-order case but denying the need for such respect in the second.¹⁶ I will try to demonstrate the untenability of the tension with an illustrative example and an explanation of why we respond (or should respond) to the example in the way I suggest.

It is relatively uncontroversial that many different sets of possible actions, protected against direct interference, could constitute an adequate sphere of concern.¹⁷ One has no complaint on autonomy grounds simply because the content of her adequate sphere includes a particular range of possible actions, relationships, conventionally defined roles, life patterns, and so on, where it could, if others had acted differently, have included some other, adequate range of actions and

¹⁵ Most second-order claims are not also first-order claims, because they are typically claims to share authority equally with others, and so are not claims to determine actions without interference. I say more about this distributive feature of second-order claims in Section III.E.

¹⁶ This may be an example of what Niko Kolodny (2010: 47) calls argument by “resonance”: “one has reason to respond to X in a way that is similar to the way that one has reason to respond to its counterpart in another dimension of importance, but that reflects the distinctive dimension to which X belongs.” Here, X is a person’s attempt to exert authority over how her sphere of concern is determined; the counterpart is her first-order action within that sphere; and the dimensions of importance are second- and first-order autonomy, respectively.

¹⁷ This is common ground of Joseph Raz, Arthur Ripstein, and Philip Pettit, for example, despite otherwise considerable differences in their arguments’ theoretical architecture.

relationships and so on. The particular content of a person's sphere of concern is largely determined by the actions of others (including long-dead others who contributed to the particular history of a society and the world in general). But this in itself does not threaten autonomy's adequacy requirement. Indeed, a demand that others should act such that my sphere of concern be adequate in such-and-such a way would be an affront to the autonomy of others—even if submitting to my demand left all others with an adequate sphere of concern.

Consider now a liberal autocrat, who claims sole authority over common life, but is firmly committed to exercising that authority in ways that respect citizens' first-order autonomy, including their claims to an adequate sphere of concern. If we have no second-order autonomy claims, there is, by hypothesis, no autonomy-based objection to this autocrat's rule. But in my view, the autocrat infringes other citizens' autonomy by denying them any authority over how the particular content of their sphere of concern is determined.¹⁸ This denial of authority denies them meaningful self-direction, or what Joseph Raz (1986: 155) calls "part authorship of one's life." Autonomy can only ground claims to part, rather than complete, authorship of our lives, because we must allow for some shaping of our lives by autonomous others. But autocratic determination of the context of our choice unduly restricts the parts of authorship to which we have a claim. Or so I will argue.

C. Alternative Explanations: Constitutionalism, Anarchism, Egalitarianism

Before I explain why the autocrat infringes citizens' autonomy despite his (first-order) liberalism, I will address some alternative explanations of negative reactions to this case. First, one might object that autocracy is a

¹⁸ Ronald Dworkin suggests a similar position in his account of equality of resources. Such equality requires more than equal bundles of resources: it requires that each person play an "equal role" in determining which, of the many possible sets of equal bundles, is the set distributed to each person (Dworkin 1981: 286–7). Dworkin explains this as resolving a concern about equality, but part of his concern is equality in protection from being subject to "arbitrary" outcomes in which one's choice did not play an adequate role—what I call an autonomy concern.

special case, perhaps because it is an especially clear case of citizens being “dependent” or “subject to the will of another”—that is, to the will of the autocrat.¹⁹ One might hold that other constitutional forms might avoid autonomy-infringing dependence or subjection without believing that autonomy requires equal sharing of authority. For instance, one might hold that the relevant requirements of non-subjection are satisfied by an oligarchic constitution in which almost everyone is disenfranchised, but the government observes separation of powers, the rule of offices rather than persons, and rule by law.²⁰ But while there are special objections to autocracy, citizens’ authorship over their lives is no less constrained by a legalistic, liberal junta determining their spheres of concern than by a single individual. In either case, citizens lack any authority over the actions of others that determine their contexts of choice. They are denied the self-direction involved in shaping their sphere of concern over time.

Second, one might agree that liberal autocrats or oligarchs violate autonomy but deny that the lesson is that we ought to share authority over common matters that determine citizens’ spheres of concern. One may take an anarchist or libertarian view of the distribution of authority. According to this view, the scope of spheres of concern is not determined by any political process involving shared authority. The proper spheres are determined by some “natural” (i.e., non-convention-dependent) morality, and the past exercise of individuals’ authority within their respective spheres—for instance, contracts or other conventions established by universal consent. Actions potentially interfering in more than one person’s sphere require the consent of each implicated person, because the actions must satisfy the requirement that each person is a full authority within their sphere, and so can veto any action that implicates their sphere.

¹⁹ On independence as a condition of autonomy or freedom, see Raz 1986: 372–3; Ripstein 2009: 217; Pettit 2012.

²⁰ Kant may have held this view; Ripstein (2009: 193, 195 n. 20) sometimes seems to endorse it. Elsewhere Ripstein (2009: 213 n. 54) seems to assume that electoral representation is necessary to satisfy Kantian requirements, though it is not clear why. Raz (1986: 156–7) similarly suggests that government coercion will not constitute invasion of autonomy if (and only if?) it guarantees rights to (democratic?) “political participation,” though, again, it is not clear why. I aim to explain the reasoning behind these suggestions. See also note 8 on the difficulty with Pettit’s argument for democracy.

An autonomy-based argument for democracy requires rejecting this anarchist-libertarian interpretation of the demands of autonomy in matters implicating more than one person. The view I have sketched is arguably more radical than the libertarianism of Robert Nozick (1974), to the extent that it, unlike Nozick's view, is unyielding in requiring a universal veto in all cases implicating more than one person's sphere of authority.²¹ Nevertheless, anarchist-libertarian positions of this kind are influential autonomy-based conceptions of justice, and so are important competitors to an autonomy-based argument for democracy.

One problem with this anarchist-libertarian view of natural spheres of individual concern is that the precise boundaries of each individual's sphere are indeterminate. Even according to the best, most fully developed natural morality (whatever that is), there will be no determinate truth about just where one individual's sphere of concern ends and another begins. (The classic case involves the boundaries of originally acquired private property in land [Stilz 2009: ch. 2; Ripstein 2009: ch. 6; Rousseau 1997a: 172].²²) Absent well-defined social conventions, the scope of spheres of concern will both be objectively indeterminate—there is no fact of the matter where the “natural” property boundary lies, for instance—and subjectively inaccessible—people will not know, and will inevitably disagree about, what the boundaries are or ought to be. This is not just a serious problem of “inconveniences” and potential conflict, as Locke put it; people enjoy less autonomy when they are uncertain about their domain of choice. Not only is there a risk that others will misjudge the boundaries (perhaps non-culpably) and invade one's sphere; agents are also faced with a dilemma regarding their own actions. Either they limit their own actions to those clearly within their own sphere of action (or those actions that gain the consent of any others whose spheres of concern might be implicated), thus dramatically

²¹ Whether or not this is true of Nozick's view depends on how you interpret his positions on, *inter alia*, the rights of the protective association vis-à-vis holdouts in the state of nature; trivial trespass; and what he calls the “shadow of the proviso” (regulating original acquisition) in affecting property rights. See Nozick 1974: chs. 4–5 and p. 180.

²² Even the meaning of control over the body raises perplexing questions, for instance involving the scope of free movement or the permissibility of trivial, unintentional bodily contact (e.g., my emitting harmless particles, too small for you to notice, that land on you a mile away).

restricting the scope of their autonomy, or they act within uncertain or indeterminate spheres, thus potentially invading the autonomy of others. The indeterminacy of natural spheres of concern renders autonomy limited and fragile in the absence of the kind of social coordination brought about, for instance, by politically generated norms.²³

The indeterminacy problem points to a more general dilemma facing the anarchist-libertarian interpretation of the demands of normative autonomy. The dilemma involves how to characterize what constitutes interference into one's sphere of concern (and thus what activity of others is subject to one's veto). On one horn, natural morality initially determines broad spheres of concern, which encompass a wide range of choices. In this case, almost anything one does will change or shape another's sphere, as one's actions block some of another's choices. But then almost any action is subject to another's veto, and it turns out the spheres of concern are inadequately narrow after all. On the other horn, spheres of concern are naturally determined to be small enough that one can act reliably without impinging on another's sphere and triggering veto power. But spheres that are so small as to be so insulated from others' actions are likely to be inadequate.

This dilemma is a highly general version of familiar problems of transactions costs and holdouts that afflict efforts at coordination among many people with veto rights (for instance, groups of property owners in the absence of public taking powers). Cooperation—even cooperation that would expand everyone's sphere of autonomy, such as public roads enabling free movement—is very difficult when everyone possesses a veto over such cooperation. But the problem is not just that procedures requiring unanimity tend to produce bad, autonomy-denying outcomes. On either horn of the dilemma, spheres of concern

²³ I do not say that such coordination requires political authority—the capacity for some agent (such as the state) to place others under obligations to obey. Collectively generated norms could perhaps serve autonomy protecting and respecting functions without any agent who generates the norms having political authority in this sense. My subsequent argument that respect for autonomy requires shared authority over the generation of these norms applies whether or not those norms are established by an entity or entities with the capacity to place others under an obligation to obey. That is, respecting equally the authority of fellow citizens may not always be equivalent to respecting the authority of a state democratically authored by those citizens. Cf. Viehoff 2014; Kolodny 2014b.

are constitutively inadequate. Respect for normative autonomy, then—which includes respecting people’s interests in the scope of their autonomous choice—requires at least some autonomy-expanding social cooperation beyond whatever is necessary to establish determinate spheres of individual concern.²⁴ And the question arises how to share authority over such cooperation.

The anarchist-libertarian approach explicitly allows that others may, individually or in aggregate, substantially shape one’s own sphere of concern without any respect for one’s own authority over that shaping. When the nature of each individual’s sphere of concern is determined by the aggregate side effects of others’ (rights-respecting) choices, each individual is denied such authority.²⁵ This approach denies agents’ claims to self-direction in the shaping of one’s sphere of concern. If that is an autonomy-based concern, the anarchist-libertarian approach is inconsistent with respect for autonomy.²⁶

Challenging my use of the liberal autocrat example from another direction, one might assert that the problem in the case is not connected to any autonomy concern, but is fully explained by the objectionable social hierarchy manifested in autocratic rule (Kolodny 2019). I agree that political inequality constitutes objectionable social hierarchy. I am

²⁴ I believe this point is vindicated even by Nozick’s work. As mentioned in note 21, he does not insist on property rules (i.e., individual vetoes) in all cases; instead, in some vital cases of social cooperation, Nozick admits a liability rule, according to which holdouts can be forced to join and then compensated.

²⁵ Compare Pettit, who argues that to constitute an (especially freedom-threatening) “invasion” of an agent’s choice, some act of interference must be triggered by the agent’s exercise of choice. Pettit allows that side effects of others’ actions (effects of actions that are not triggered specifically in response to one’s own choice) may “vitalize” freedom when they decrease one’s resources for exercising choice (Pettit 2012: ch. 1). But vitiation only constitutes lack of freedom when resources fall below a basic minimum (Pettit 2012: ch. 2). I believe this is a gap in his view—a gap that leaves the view less able to identify threats to freedom from non-agential structures that limit self-direction even in the absence of agential invasion (or capacity for invasion). It is not that Pettit is wrong to say that in these cases we should often let the liberties and their effects lie. It is just that the effects, in my view, trigger autonomy-based claims to authority over the actions that produce the effects.

²⁶ In my view, this is the best explanation of the Kantian claim that we wrong another when we “unilaterally” claim rights (e.g., to property) in a state of nature. (See, e.g., Ripstein 2009: ch. 6.) Like Locke, and unlike Ripstein’s Kant, I believe it is permissible to defend one’s rights forcibly outside of a “lawful” condition. But we nevertheless do not fully respect one another’s autonomy if we do not organize our exercise of rights in a way that respects each person’s authority over the ways that exercise shapes their sphere of concern. A state or state-like institution (note 23) may be the most reliable way to guarantee such respect.

also open to the suggestion that the objectionable features of the case do not depend only on citizens' being "under the power" of others, since there may be an equivalent sense in which citizens of a democracy are under the power of officials (or each other). As I will further argue shortly, I believe second-order autonomy can be infringed unintentionally, as the side effect of many separate actions, a case not involving being under the power of these several others in any straightforward sense. Still, I think it is a mistake to believe that a concern about social equality, *independent of any concern of autonomy*, fully explains our concerns about the autocrat and related cases. Liberal autocracy would be objectionable even if the autocrat foreswore any economic gain or general social precedence or any other constituent of social inequality (beyond the authority to direct social interaction so as to preserve first-order autonomy). Of course, if we all have claims to equally share authority given second-order autonomy claims, denying the claims of some constitutes a serious, socially salient inequality. But the seriousness of the inequality would be at least partly explained by the autonomy claim, so the former would not serve as a way to explain away the latter.²⁷

D. Direction and Implication of the Will

If these alternative explanations for disquiet about liberal autocracy fail, what is the right explanation? The above reflections suggest two lessons. First, indirect effects of actions (often in aggregate) can trigger autonomy claims by threatening agents' self-direction. This is an important correction to Kantian, republican, and anarchic approaches, which, in their various ways, focus on agent-to-agent invasion or subjection. Second, and relatedly, there is an infringement of self-direction not only when one's first-order autonomy is violated, but also when one is denied authority over how others shape the sphere of concern within which

²⁷ Institutional consequences follow: the success of the autonomy argument would deny the claim of some egalitarians, such as Niko Kolodny (2014b) and Daniel Viehoff (2014), that political equality is compatible with "equal-but-zero" political power, such as by random selection of laws. (See also David Estlund's [2008: ch. 4] claim that this result is something of a *reductio ad absurdum* for procedural defenses of democracy.)

one exercises first-order autonomy. This is an important correction to liberal aristocracy. I believe we can best explain these lessons by saying that we have autonomy claims to authority over how others *direct* or *implicate* our wills.

Other people's actions can put great pressure on how I can—and how I will—act within any given sphere of concern. In this sense they *direct* my will (intentionally or not). This pressure can come in the form of limitations, physical or social. Recall the example of neighbors building walls (on their own property) around another while she sleeps. While some kinds of limitations may be so extreme as to violate one's first-order autonomy claims, other limits may be less severe, but nevertheless substantially shape the scope of one's choice. Pressure on one's choice can also come in the form of imposing obligations that constrain the exercise of that choice. For instance, I will find it difficult to autonomously exercise my religion if I must spend all my free time saving others who fall into the stream that borders my land (or if I must spend time exercising due care in posting warnings). Even if we make no assumptions about strong obligations of justice, minimal social interaction typically brings with it obligations (such as duties to engage in easy rescue) that constrain one's choices.²⁸ The same interests in self-direction that underlie our first-order autonomy claims—interests not only in having our own psychological states determine our choice, but also in having sufficient external scope for self-direction—also support second-order claims to authority over how others will shape one's context of choice.

The liberal autocrat accepts that others shape one's sphere of concern, but argues that this only implies that government should organize social interaction in whatever way best protects first-order autonomy claims. The government should regulate wall-building and define obligations to put up warning signs by perilous streams and so on. But this is just to replace one way of shaping our spheres of concern with another. By

²⁸ Even those who doubt others can impose many duties on us should agree that the actions of others can put substantial pressure on our deliberations even if they do not generate obligations. For example, danger invites non-obligatory rescue.

hypothesis, it is a good way, in terms of responding to the first-order claims of normative autonomy. But normative autonomy in general is not about producing the best outcomes—not even the most autonomy-producing outcomes. An adult's first-order autonomy claims are not generally respected by forcing him to engage in a program of education that will lead him to choose more wisely and authentically, or to choose in ways that will best preserve a large sphere of choice throughout his life. Instead, autonomy claims are claims to direct one's choice. Similarly, we should understand second-order autonomy claims not as a claim that others shape our context of choice in whatever way best satisfies our first-order claims, but as a claim to direct that shaping ourselves (along with others who have equal autonomy claims with respect to that context). Others, by shaping our spheres of concern, direct our wills, whether they mean to do so or not. Respecting autonomy requires granting individuals authority over how others engage in that direction.

The second-order autonomy concern about others directing our wills focuses on the autonomy claimant as a patient—as someone affected by the will-directing actions of others. There are other grounds for second-order autonomy claims we can recognize when considering the autonomy claimant as an agent. Many of our actions affect others. We generally have responsibility for how our actions affect others. Our actions affect others in particular ways because of the context, substantially shaped by others, in which they take place. Our responsibility thus takes the particular form it does in significant part because of the context-shaping actions of others. This, I now argue, gives us an autonomy claim regarding how those others shape our responsibility.

Autonomy is closely connected to responsibility. Arguably, we are only fit to be held responsible to the extent we are autonomous, or at least to the extent we are fit to be treated as autonomous. Part of what it is to treat others as autonomous is to hold them responsible for their actions in various ways; part of what it is to understand oneself as autonomous is to hold oneself responsible in various ways, including by understanding one's actions as one's own, and understanding oneself as connected to at least some of the consequences of one's actions in certain important ways. Without attempting to analyze fully the nature of our responsibility, I will summarize these connections by saying that

autonomous agents are *implicated* in their actions. That is, they bear these kinds of responsibility for their actions. Autonomous agents have a deep interest in the nature of this implication, and this interest is closely connected to the exercise of the agent's autonomy itself.

Return to the fact that the actions of others shape the context of our choice, and thus the meaning and effects of our actions. Accordingly, these others shape how exactly we are implicated in our actions, and what this implication involves. Other agents shape the nature of our responsibility. To take a political example, one's actions often contribute (often unintentionally) to the maintenance of some social order—whether a very local order such as a family structure, or a more global order such as a political regime, or to many orders at once. It can sometimes be very difficult to disavow or prevent this contribution entirely. Perhaps this difficulty renders the contribution permissible, even if the prevailing order is unjust.²⁹ But this permissibility does not always (or even usually) absolve one entirely of responsibility for the contribution. (The contribution may be permissible for reasons that do not derive from extreme, responsibility-negating duress.) Even in cases in which we have no responsibility for the contribution in the sense of liability to moral censure (blame, resentment, shame, punishment, etc.), we may have some other form of responsibility stemming from the way in which we autonomously chose the contributing action and helped produce certain outcomes. This responsibility may be serious even if it does not ground blame—it could, for instance, entail duties to compensate (which sometimes operate on strict liability principles, according to which one is liable even if not blameworthy), or merely agent regret—a form of regret for contributing to a bad outcome, even if blamelessly. These forms of responsibility are meaningful and potentially substantial, and the precise shape they take stems from how others shape our context of choice. Others thereby implicate our wills—they shape the ways in which we will exercise and experience responsibility—in ways that go to

²⁹ For example, active disavowal may be too costly to oneself, and thus not required; disavowal may not meaningfully change anything for the better, and thus not be required; disavowal may be too costly for others, and thus prohibited, even if the current order is unjust.

the core of our exercise of autonomy. Our autonomy is infringed if we have no ability to direct how others implicate our wills. Accordingly, I believe, we have second-order autonomy claims to authority over how our wills will be so implicated.³⁰

Social interaction, then, both directs and implicates our wills. It shapes and constrains our choice, and it shapes our responsibility and our relationship to our own actions. Autonomous agents have claims to authority over how their wills are directed and implicated. This explains our two lessons: that autonomy has a second-order dimension, and that this dimension includes concerns about the aggregate effects of others' actions, and not only intentional invasions of autonomy. When social interaction directs or implicates the wills of many people, each of those people has a claim to authority over that interaction. Because the universal-veto model of equalizing authority fails to respect autonomy claims, simultaneously respecting everyone's claims must occur on some other basis. The only other satisfactory basis, if we respect people as moral equals, is to share that authority on equal, democratic terms. Thus there is a strong autonomy-based reason to organize social life democratically.³¹

³⁰ Thomas Nagel (2005: 128, 129) rightly emphasizes that the state's "engagement of the will" of citizens renders them responsible in some way for state action (whether or not the state is democratic). While I believe Nagel (2005: 129 n. 14) is wrong to suggest that this engagement of citizens' wills depends on whether the state explicitly acts "in the name" of citizens, he is right to insist that certain obligations are triggered not only by acting on one another as patients, but also by acting in ways that implicate the agency of those we act on or with.

³¹ Does this argument support *democracy*, or simply the initially equal distribution of claims to authority that citizens might autonomously alienate (e.g., by selling)? The question deserves more discussion, but here are a few doubts about the alienability of second-order autonomy claims. (1) If one's claims are to authority over all directing and implicating actions of others, there are questions whether bilateral contracts (e.g., selling votes) would extinguish the claims of the seller vis-à-vis everyone other than the buyer. We would face questions about whether such contracts could obviate the autonomy-respecting duties of third parties. (2) Perhaps autonomy requires sustained authority over time, which is incompatible with many forms of long-term alienation, along the lines of arguments against slave contracts. So alienation of some incidents of second-order autonomy might be permissible, but not wholesale alienation of core rights that constitute democratic citizenship. (3) There may be non-autonomy-based reasons to think such alienation void (perhaps sounding in equality), though I grant that relying on such reasons alone might introduce tensions between different arguments for democracy. I thank Chiara Cordelli for raising this question.

E. Democracy and Self-Government Revisited

It is virtually impossible, given substantial social interaction, to avoid directing and implicating the wills of others. This remains true even in the most democratic regimes. Citizens in democracies regularly direct and implicate the wills of others, not just through the exercise of formal political power, but also through their everyday actions, individually and in aggregate. There need be no autonomy violation in this, however.³² One's second-order autonomy claims are not to preventing all direction and implication by others. Instead, they are claims to authority over the direction and implication to which they are subject. This is an ideal of autonomy neither as unilateral control nor as freedom from rule. Instead, it is an ideal of multilateral, mutual shaping of wills on equal terms, given the inevitability of some such shaping. So the fact that mutual direction and implication of the will persists in democracies is no embarrassment to the argument.

This framing may, however, reinforce concerns that this ideal of autonomy is too thin or weak. The concern, recalling from Section I, is that democracy does not guarantee to citizens that their judgments will actually determine collective actions—not even collective actions that regulate the citizens' own actions. Therefore, the thought continues, this cannot be a meaningful form of self-government or a way to respect citizen autonomy.

Second-order autonomy claims are not claims to actual determination of collective choices—or, more generally, to actual determination of one's sphere of concern. But this does not mean that the satisfaction of these claims is inconsequential. On the contrary, when one's authority is recognized, others accept and act upon obligations to respond positively in their deliberations to one's judgments about what to do. This is a real social effect of one's own choice. In a democratic society, this fact of mutual recognition of authority is, moreover, publicly visible in the organization of collective decision-making.

³² Some actions, including democratic laws, might infringe on some citizen's first-order autonomy. But directing and implicating another's will is not itself an infringement.

The reason that this share of authority—consideration as an equal—does not amount to actual determination of outcomes is that each citizen has a symmetric claim to such authority. There is a distributive concern that shapes how autonomy interests are translated into decisive claims on how others ought to act. An interest in self-direction sometimes grounds near unilateral control over some choice—that is, first-order autonomy, such as one’s unilateral right to eat one’s apple without the permission of others. Even here, conclusions that one is entitled to such unilateral control depend on distributive concerns about what exercises of free choice are suitably compatible with the freedom of others.³³ In other cases, the distributional concern is more constraining on individuals, because for a given choice—such as how to set the basic terms of social interaction—each individual has an equally strong autonomy interest, grounding a claim to equal authority. So when it comes to second-order autonomy, one’s claim is to equal, mutual authority rather than unilateral authority. When this claim is honored, this constitutes a real and meaningful response to one’s autonomy, even though it does not amount to individual determination of collective action.

IV. How Much Interaction Is Enough?

Social interaction of a suitably dense and significant kind directs and implicates the wills of participants in ways that trigger second-order autonomy claims to share authority over that interaction democratically. But how dense and significant does interaction have to be to trigger autonomy claims? What, exactly, does it take to direct or implicate wills in the relevant way?

Absent specified limits, the logic of the argument might imply that any interaction that has any effect on anyone triggers the affected person’s second-order autonomy claims. Some have defended an “all-affected interests principle,” which requires that anyone affected by some policy

³³ The distributional concern here is not about the fair distribution of goods, but the fair or proper assignment of moral claims, given the autonomy interests of potentially affected agents. See Wallace 2019.

has the right to participate in deciding whether to undertake the policy (Goodin 2007; cf. Beitz 2011). Typically, defenders of the principle do not provide a deep justification for the principle itself; the autonomy-based argument I have provided could serve as a foundation for the principle. Some, however, might think it a problem for the autonomy argument that it endorses the all-affected principle, because the principle seems to result in the supposedly absurd conclusion that virtually every action anyone engages in must be subject to democratic procedures. The objection generates nightmare scenarios of suitors voting on whom the beloved will marry, or of everyone in an agricultural supply chain deciding on what you eat for lunch.³⁴

While I do not think millions of people should determine by vote the contents of your next meal, I believe most criticisms of the all-affected principle are overwrought, because based on a misunderstanding of what the principle actually requires (Wilson unpublished b).³⁵ Nevertheless, my present aim does not require defending such an ambitious principle. It is enough for now to argue that political societies as we have historically experienced them should be organized democratically. For that purpose, we can remain provisionally agnostic on questions about how the borders of societies should be drawn, the rights of affected foreigners to participate in decision-making, and so on.³⁶ Instead, we can focus on whether societies of a relatively familiar kind direct and implicate the wills of members in ways sufficient to trigger second-order autonomy claims, and the attendant claims to democracy.

³⁴ Niko Kolodny (2014a: 222–3) states that this was his reason for abandoning his earlier (unpublished) defense of the claim that people have an irreducible interest in having a say over what affects their sphere of concern. The marriage example comes from Nozick (1974: 269).

³⁵ For the curious or concerned: I argue (Wilson unpublished b) that the all-affected principle should be interpreted to ground authority over the terms of social interaction (e.g., laws), rather than every individual decision. These laws may (and ought to) allow space for individual autonomy. The principle does require a much more democratic transnational order, but does not itself require a one-person-one-vote global democracy. The principle has critical bite but avoids the most concerning implications and supposed paradoxes.

³⁶ A complete defense of the argument would require asking what the implications of the argument are for questions such as these, and ensuring that those implications were suitably defensible, thereby ensuring that the argument survives in reflective equilibrium. But I will consider the argument successful enough for the present if it can deliver the implication that there are strong autonomy-based reasons for societies of a relatively familiar kind to be democratic.

Here is one route to the conclusion that they do. Say that a person is in a structural position within a society if that person occupies a social role defined by relatively narrow social expectations regarding a wide range of behaviors—expectations that carry with them consequences for performance or violation, including liability to certain reactive attitudes such as praise or blame. (People may occupy more than one structural position at once.) Say that two structural positions are interdependent to the extent that the actions of those in one position affect the expectations, or the opportunities to satisfy the expectations, attaching to those in the other position. When a group contains people who occupy structural positions organized in a network of highly interdependent positions, say that a basic structure exists for the society comprising that group of people.³⁷ When a basic structure in this sense exists, there are good reasons to think that almost any pattern of activities will direct and/or implicate the wills of many others (perhaps all others) in the society. So everyone's second-order autonomy claims to authority over the pattern of activities are triggered by any pattern, or proposed change of pattern. In that case, everyone has a claim to a political institution that shares authority democratically over the norms that will regulate the patterns of activity—typically, the laws.

I submit that societies are generally characterized by basic structures in the above sense. Therefore members of those societies have claims to democratic authority over social life. This argument leaves open whether second-order autonomy claims to democratic authority extend beyond the conventional boundaries of societies, and what, within a given society, is the appropriate scope of democratic decision-making. I believe the argument does establish, however, an autonomy-based argument for democracy in such societies.

³⁷ I am influenced here by A.J. Julius. Unlike Julius (2003: 328), I do not rely on the concept of "framing"—that is, acting with the intention of leading another to act in a way that advances one's own interest. Others may implicate and direct one's will without framing one (perhaps because the effects on one are unintentional, or not intended to advance the actor's interest). Conversely, it may be possible that framing actions may not actually direct or implicate the wills of others (perhaps because the action fails to meaningfully affect the choice of the other). Relatedly, I do not claim that implication or direction of the will triggers claims to egalitarian distributive justice—only claims to share authority.

Conclusion

Democracy has an essential connection with freedom of a kind. It is not just that democratic societies tend to protect and celebrate certain individual freedoms. Democracies respect the autonomy of their citizens by respecting their claim to authority over patterns of social interaction that in turn direct and implicate the wills of those citizens. Democracy does not guarantee anyone control over her social world. But the proper autonomy-based argument for democracy is not a demand for such chimerical control; it is a demand for the co-satisfaction of every person's claims to autonomy. This does not amount to obeying only laws one gives oneself, as Rousseau would put it, since each individual does not give herself all laws to which she is subject. But democratic citizens observe a related ideal, that of engaging in choice subject only to those social constraints that are subject to our own authority, as much as is possible given the equal authority claims of others.

This is only one part of what is required to respect the autonomy of persons. Such respect also requires more familiar protections for individuals directing their own lives within certain bounds—respecting first-order autonomy claims. Societies must work to respect both types of demands of normative autonomy, protecting first-order autonomy through democratic means that respect second-order autonomy. These aims may sometimes be in tension. But this does not mean democracy and autonomy are essentially at odds; instead, it means that two dimensions or aspects of autonomy require active coordination.

This point mutes the sting of Christiano's argument, noted in Section I, that democratic government reduces self-government (by collectively passing laws) more than it empowers self-government (by granting individual political rights). There is no inconsistency in claiming that both first- and second-order autonomy demand respect, as both can in principle be satisfied simultaneously. Moreover, the absence of collective government does not amount to unlimited individual self-government. Instead, it leaves everyone to direct and implicate the wills of others without regard for those others' autonomy claims, and leaves most people

with unduly limited spheres of concern. Finally, even if many aspects of private or first-order self-government are more valuable than public or second-order self-government, fully respecting both orders of self-government is necessary to satisfy each individual's autonomy claims.

Working out how to simultaneously respect all citizens' claims to both first- and second-order autonomy is a great challenge of social life. The challenge is urgent because democracy is part of the realization of freedom as autonomy, not merely a possible and uncertain means to that freedom.³⁸ This constitutive connection between democracy and autonomy may be one part of the reason democratic societies generally respect citizens' autonomy, at least relative to non-democratic societies: to be disposed to democracy is to be disposed to respect autonomy in a certain way, and vice-versa.

This autonomy-based argument for democracy provides us reasons of considerable strength to organize social life democratically. It also, I believe, supports egalitarian arguments for democracy. Democracy is valuable in part because it constitutes a form of egalitarian relation between citizens. The autonomy argument helps explain why this is an important and valuable kind of equality—and, moreover, why it is that political equality is an essential part of more general social equality. It is because political equality involves the equal sharing of authority, and thus equal respect for citizens' autonomy. It is precisely because claims to this respect are of great moment that equally recognizing those claims is such a defining feature of social relations.³⁹

³⁸ In this I agree with those who claim that arguments for democracy and for individual rights share common content or structure (Waldron 1999; Brettschneider 2007).

³⁹ Thanks to participants at the Oxford Studies in Political Philosophy Workshop for valuable questions, and especially to Anna Stilz for insightful comments. Thanks also to participants at the Yale Political Theory Workshop, especially to Hélène Landemore and Bryan Garston. For helpful discussions of an early draft of this chapter, I thank Eric Beerbohm, Mark Berger, Chiara Cordelli, Max Etchemendy, Ben Laurence, Cristian Pérez Muñoz, Julie Rose, and Emma Saunders-Hastings. For thoughtful, detailed comments on a recent draft, I thank Chiara Cordelli (again) and Shmulik Nili. I am grateful to the editors of *Oxford Studies in Political Philosophy* and two anonymous reviewers for helpful suggestions at the final stages of revision.

References

- Aristotle. 1984. *The Politics*. Chicago, IL: University of Chicago Press.
- Arneson, Richard J. 2004. "Democracy Is Not Intrinsically Just." In *Justice and Democracy: Essays for Brian Barry*. Cambridge: Cambridge University Press, 40–58.
- Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.
- Beitz, Charles R. 2011. "Global Political Justice and the 'Democratic Deficit.'" In Wallace et al., eds., *Reasons and Recognition: Essays on the Philosophy of T.M. Scanlon*. New York: Oxford University Press, 231–55.
- Brettschneider, Corey. 2007. *Democratic Rights: The Substance of Self-Government*. Princeton, NJ: Princeton University Press.
- Buss, Sarah. 2005. "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints." *Ethics* 115 (2): 195–235.
- Christiano, Thomas. 1996. *The Rule of the Many: Fundamental Issues in Democratic Theory*. Boulder, CO: Westview.
- Christiano, Thomas. 2011. "An Instrumental Argument for a Human Right to Democracy." *Philosophy & Public Affairs* 39 (2): 142–76.
- Darwall, Stephen. 2006. "The Value of Autonomy and the Autonomy of the Will." *Ethics* 116 (2): 263–84.
- Dworkin, Ronald. 1981. "What Is Equality? Part 2: Equality of Resources." *Philosophy & Public Affairs* 10 (4): 283–345.
- Estlund, David. 2008. *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.
- Gilbert, Margaret. 1996. *Living Together*. London: Rowman and Littlefield.
- Goodin, Robert. 2007. "Enfranchising All Affected Interests, and Its Alternatives." *Philosophy & Public Affairs* 35 (1): 40–68.
- Hobbes, Thomas. 1994. *Leviathan*. Ed. Edwin Curley. Indianapolis, IN: Hackett.
- Julius, A.J. 2003. "Basic Structure and the Value of Equality." *Philosophy & Public Affairs* 31 (4): 321–55.
- Kolodny, Niko. 2010. "Which Relationships Justify Partiality? The Case of Parents and Children." *Philosophy & Public Affairs* 38 (1): 37–75.

- Kolodny, Niko. 2014a. "Rule Over None I: What Justifies Democracy?" *Philosophy & Public Affairs* 42 (3): 195–229.
- Kolodny, Niko. 2014b. "Rule Over None II: Social Equality and the Justification of Democracy." *Philosophy & Public Affairs* 42 (4): 287–336.
- Kolodny, Niko. 2019. "Being Under the Power of Others." In Elazar and Rousseliere, eds., *Republicanism and the Future of Democracy*. Cambridge: Cambridge University Press, 94–114.
- List, Christian, and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.
- Nagel, Thomas. 2005. "The Problem of Global Justice." *Philosophy & Public Affairs* 33 (2): 113–47.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Pettit, Philip. 2012. *On the People's Terms: A Republican Theory and Model of Democracy*. Cambridge: Cambridge University Press.
- Plato. 1991. *Republic*. Trans. Allan Bloom. 2nd ed. New York: Basic Books.
- Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.
- Ripstein, Arthur. 2006. "Beyond the Harm Principle." *Philosophy & Public Affairs* 34 (3): 216–46.
- Ripstein, Arthur. 2009. *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press.
- Rousseau, Jean-Jacques. 1997a. *A Discourse on the Origin of Inequality*. In *The Discourses and Other Early Political Writings*. Trans. Victor Gourevitch. Cambridge: Cambridge University Press, 114–222.
- Rousseau, Jean-Jacques. 1997b. *On the Social Contract*. In *The Social Contract and Other Later Political Writings*. Trans. Victor Gourevitch. Cambridge: Cambridge University Press, 40–152.
- Shiffrin, Seana Valentine. 2000. "Paternalism, Unconscionability Doctrine, and Accommodation." *Philosophy & Public Affairs* 29 (3): 205–50.
- Stilz, Anna. 2009. *Liberal Loyalty: Freedom, Obligation and the State*. Princeton, NJ: Princeton University Press.
- Stilz, Anna. 2019. *Territorial Sovereignty: A Philosophical Exploration*. Oxford: Oxford University Press.
- Tuck, Richard. 2008. *Free Riding*. Cambridge, MA: Harvard University Press.
- Viehoff, Daniel. 2014. "Democratic Equality and Political Authority." *Philosophy & Public Affairs* 42 (4): 337–75.

- Waldron, Jeremy. 1999. "Participation: The Right of Rights." In *Law and Disagreement*. Oxford: Oxford University Press, 232–54.
- Wallace, R. Jay. 2019. *The Moral Nexus*. Princeton, NJ: Princeton University Press.
- Wilson, James Lindley. 2019. *Democratic Equality*. Princeton, NJ: Princeton University Press.
- Wilson, James Lindley. Unpublished a. "Against Instrumentalism about Democracy." Unpublished manuscript.
- Wilson, James Lindley. Unpublished b. "Making the All-Affected Principle Safe for Democracy." Unpublished manuscript.

8

Why Strict Compliance?

Simon Căbulea May

Introduction

John Rawls (1971/1999) defends the principles of justice as fairness as the appropriate terms of association for a well-ordered society. He presents this defense as an exercise in ideal theory (1971/1999, 8). Ideal theory has two key features in his account. First, it rests on a strict compliance assumption: the institutions of a well-ordered society are perfectly just and every person acts as justice requires in upholding these institutions. Second, ideal theory takes priority over non-ideal theory insofar as it underpins the moral norms that govern responses to injustice in societies that are not well ordered.¹ The combination of these two features creates a puzzle: why should moral norms for responding to the widespread injustice of the actual world depend on an account of a hypothetical world in which every agent complies with the requirements of justice? Rawls's approach may seem excessively perfectionist. Solutions to the urgent problems of everyday life do not generally rely on appeals to any standard of perfection. Moral arguments for particular tax codes, highway speed limits, and workplace safety regulations, for instance, should not presume that every relevant agent will comply with the rules once they are implemented. So why should claims about how to build a more

¹ Rawls also defines ideal theory in terms of an assumption of favorable circumstances (1971/1999, 216). Although this is an important element of Rawlsian ideal theory, my present argument concerns only the strict compliance assumption. I do not discuss alternative concepts of ideal theory that do not employ a strict compliance assumption. See Valentini (2012) for discussion of various accounts of ideal theory. Nor do I discuss the full array of issues associated with Rawlsian ideal theory. For a comprehensive discussion of these issues, see Stemplowska and Swift (2014).

just world ultimately depend on any assumption of strict compliance in a perfectly just society?

In this chapter, I present a distinctly Rawlsian interpretation of ideal theory, one that provides a response to the strict compliance puzzle.² This *cooperative interpretation* is grounded in the idea of society as a fair scheme of cooperation, which Rawls (2001, 5) describes as the most fundamental idea of justice as fairness. A key element of the Rawlsian idea of cooperation, I claim, is that the individual participants of a genuinely cooperative scheme—whatever its scale—are morally accountable to each other for complying with the scheme's rules. This means that each participant has the moral standing to demand of the others that they comply with the rules. I argue that the logic of these moral demands requires that the scheme's rules be worked out on the basis of a strict compliance assumption. The ideal theory of cooperation takes priority over non-ideal theory because responses to unfairness presuppose some conception of fair cooperation between agents who comply with the rules of the scheme. In justice as fairness, society as a whole is a grand scheme of cooperation. The principles of justice constitute the moral terms of association for this cooperative scheme, and hence define the moral demands that citizens, as such, may make of one another. Thus, these principles of justice must likewise be worked out on the basis of a strict compliance assumption. Once the principles of justice have been established, moral norms for responding to injustice can be developed.

The cooperative interpretation is *deontic* insofar as it is grounded in the way that the participants of a genuinely cooperative scheme are obligated to follow its rules.³ The most obvious interpretation suggested by Rawls's own statement of justice as fairness, in contrast, is *telic*. On this standard interpretation, the function of ideal theory is to specify a political telos—a realistic utopia—that should orient political decision

² The interpretation is distinctly Rawlsian in the sense that it depends on certain basic ideas that are either an explicit part of or closely related to Rawls's broader statement of justice as fairness. I do not claim that the interpretation is in fact Rawls's own view of ideal theory.

³ The cooperative interpretation is only one possible deontic interpretation of ideal theory. Blain Neufeld (2017) argues that ideal theory is required by a principle of equal civic respect, a principle that underpins the idea of public reason. Despite some important similarities with Neufeld's view, the cooperative interpretation does not in itself support the argument for public reason. See also Laden (2013).

making in the actual world. Ideal theory takes priority over non-ideal theory because it identifies the ultimate end to be realized, and this end must be identified before the appropriate means to it can be determined. Ingrid Robeyns aptly captures the spirit of this telic interpretation:

Ideal theory functions as a mythical *Paradise Island*. We have heard wonderful stories about Paradise Island, but no one has ever visited it, and some doubt it truly exists. We have a few maps that tell us, roughly, where it should be situated, but since it is in the middle of the ocean, far away from all known societies, no one knows *precisely* where it is situated. Yet we dream of going there, and ask ourselves how we could get there, and in which direction we should be moving in order to eventually reach Paradise Island. . . . We don't know whether it can be reached and no one has ever set foot on the Island. Yet since it is our dream to go there, reaching Paradise Island is our ultimate goal.

(2008, 344–45)

My aim in this chapter is not to criticize the telic interpretation, but rather to establish the cooperative approach as a viable alternative response to the strict compliance puzzle. On the cooperative interpretation, the idea of a well-ordered society in justice as fairness may play a telic role, but only because it is already a quite specific substantive ideal of social cooperation. The strict compliance assumption is ultimately warranted, not because the Rawlsian well-ordered society serves as the telos for political decision making, but because this society is a grand scheme of fair cooperation between citizens who have standing to press demands of justice against one another.

The chapter contains six sections. The first four sections present the cooperative interpretation of ideal theory. Section I outlines Rawls's cooperative conception of society. Section II presents an account of the nature of moral demands that participants in a scheme of cooperation may make of one another. Section III explains why a moral theory of cooperation, so conceived, must incorporate a strict compliance assumption. Section IV extends the argument to the political domain of justice and the idea of a well-ordered society. The last two sections contrast the cooperative interpretation with the standard telic approach. Section V

discusses an anti-utopian objection to the telic interpretation of ideal theory and explains why this objection does not threaten the cooperative alternative. Section VI explains how the cooperative interpretation allows for the possibility of a non-utopian ideal theory. A moral theory of cooperation need not assume that the cooperative scheme in question is anything more than a means to some other end. Paradise Island does not have to exist for the strict compliance assumption to be warranted. And even if the ideal theory of justice as fairness is ultimately utopian, it is not true that it is ideal because it is utopian.

I. The Cooperative Conception of Society

Rawls presents justice as fairness as an alternative to utilitarianism and draws several contrasts between the two doctrines. Some of these contrasts concern the different institutional implications the doctrines have, such as for the protection of individual liberties and the structure of a just economic system. These differences are important in determining just public policy, but the deepest philosophical contrast between the two doctrines is the difference in their underlying conceptions of society. Rawls claims that a conception of justice is part of an overall social ideal:

A social ideal in turn is connected with a conception of society, a vision of the way in which the aims and purposes of social cooperation are to be understood. The various conceptions of justice are the outgrowth of different notions of society against the backdrop of opposing views of the natural necessities and opportunities of human life. Fully to understand a conception of justice we must make explicit the conception of social cooperation from which it derives. (1971/1999, 11)

Justice as fairness and utilitarianism derive from quite different normative conceptions of society (1971/1999, 29–30). Justice as fairness conceives of a well-ordered society as a scheme of cooperation between free and equal persons for their reciprocal advantage. Utilitarianism, in contrast, conceives of a well-ordered society as one in which social interactions and resources are efficiently administered to maximize the

satisfaction of desires. Rawls's normative conception of society provides the first premise for the cooperative interpretation of ideal theory:

- (1) A just political society constitutes a grand scheme of social cooperation between free and equal persons for reciprocal advantage.

As it stands, the cooperative conception of society is a somewhat nebulous idea. Two steps help specify its meaning more determinately: a fuller account of the contrast with the utilitarian administrative conception, and an explanation of the idea of genuine social cooperation it incorporates.

1. Essential Plurality and Relational Constraints

The contrast between the cooperative and administrative conceptions of society finds expression in at least two ways. First, justice as fairness is essentially pluralistic in the sense that it requires that the content of the principles of justice depend on their role in setting the terms of association between a plurality of individuals with competing interests. Utilitarianism can allow that the concept of justice is defined by the role of the principles in determining the appropriate distribution of benefits and burdens across a group of people. So, for both doctrines, the problem of justice is essentially pluralistic. But the utilitarian solution to this problem is not. Rawls claims utilitarianism "does not take seriously the distinction between persons" because it extends the principle of rationality for a single person to society as a whole (1971/1999, 24). The utilitarian legislator regards individuals in much the same way as an entrepreneur regards her various investments: what matters is how the allocation of rights and duties to these individuals affects the bottom line; that is, maximal aggregate utility. The principle of utilitarian legislation is the same whether or not it is the interests of a single individual at stake or those of many millions of people.

Rawls's "separateness of persons" objection characterizes utilitarianism as, in one sense, insufficiently individualistic (1971/1999, 26). But Rawls also claims that the doctrine's use of the principle of rationality for

a single person demonstrates that it is, in another sense, overly individualistic:

[The] assimilation of justice to a higher order executive decision... is central to classical utilitarianism; and it also brings out its profound individualism in one sense of this ambiguous word. It regards people as so many *separate* directions in which benefits and burdens may be assigned; and the value of the satisfaction or dissatisfaction of desire is not thought to depend in any way on the moral relations in which individuals stand, or on the kinds of claims which they are willing, in the pursuit of their interests, to press on each other.

(Rawls 1958, 187; 1999a, 66)

The second contrast between the cooperative and administrative conceptions of society, then, is that the principles of justice as fairness depend on the nature of the proper moral relations between individual citizens. The idea here is not simply that principles of justice should have a certain substantive content that defines what these relations should be. On that reading, citizens' relations would be purely derivative of the principles, so there would be no reason in utilitarianism why they should constrain the value of desire satisfaction. Rather, Rawls means some account of the moral relations between citizens serves as a constraint on the principles of justice. Refer to this feature of Rawls's view as a *relational constraint*.

A relational constraint entails that the moral principles governing some sphere of social life depend on an account of the appropriate moral relations between the people in that sphere. Consider, for example, the problem of determining which moral principles govern the domestic life of a family. These principles plausibly depend on assumptions about the appropriate relations between members of the same family: parents and children have associative obligations of care toward each other that they do not have toward outsiders. Similarly, principles of family life should accommodate and legitimate the love and affection family members are normally expected to feel for each other: upsetting one's parents matters in a way that upsetting the other members of a student dormitory does not, precisely because they are one's parents and not merely

fellow residents of a shared accommodation. But justice as fairness cannot incorporate a relational constraint in the exact same way as the ethics of family life. This is because the Rawlsian well-ordered society is not a community dedicated to the pursuit of some overarching conception of the good, much less a family united by the bonds of love. There is no place in justice as fairness for the idea that relations between citizens should mirror the thick ethical relations within families or other social associations. Justice as fairness may invoke only a thin account of citizens' moral relations. The resources for this thin theory must be provided by the idea of social cooperation between free and equal persons in itself.

2. Genuine Social Cooperation

Rawls distinguishes genuine social cooperation from socially coordinated activity. Both involve schemes of social interaction, but they differ in two key respects, each of which underpins a relational constraint. The first respect is an explicit feature of justice as fairness. Rawls claims social cooperation embodies certain values, whereas mere coordinated activity need not: "The idea of cooperation includes the idea of fair terms of cooperation [and these] specify an idea of reciprocity, or mutuality: all who do their part as the recognized rules require are to benefit as specified by a public and agreed-upon standard" (2001, 6). Genuine social cooperation is fair and reciprocal, whereas efficiently coordinated activity need not be. The first relational constraint, then, is that the principles of justice are justified only if they constitute fair and reciprocal relations between the citizens of a well-ordered society. Thus, Rawls defends his difference principle in part on the grounds that it better expresses the value of reciprocity than a utilitarian alternative (1971/1999, 64, 76).⁴

⁴ See also Rawls (2001, 96): "Between [the social contract and utilitarian] traditions there is a basic contrast: the idea of society as a fair scheme of cooperation is quite naturally specified so as to include the ideas of equality...and reciprocity (of which the difference principle is an example). By contrast, the idea of society organized to produce the most good expresses a maximizing and aggregative principle of political justice. In utilitarianism, the ideas of equality

The second respect in which genuine social cooperation differs from socially coordinated activity is less explicitly developed in Rawls's statement of justice as fairness, but it is pivotal to the cooperative interpretation of ideal theory. The participants of a cooperative scheme, I claim, are mutually accountable to each other in the sense that they have standing, as participants, to make demands of one another in accordance with the rules of the scheme. The rules define not only the appropriate content of the demands but also the nature of the standing the participants have to press them. Socially coordinated activity, in contrast, does not require any such relations of mutual accountability. A central authority could impose various rules on the participants and take responsibility for enforcing those rules. The rules are sent out by the authority to each participant and it is to this authority alone that the members are accountable. In essence, its relations of accountability are fundamentally vertical, rather than horizontal.⁵

To illustrate the idea of relations of mutual accountability, consider the difference between two normative conceptions of the firm. In the first conception, the firm's workers determine the rules governing their joint labor and, crucially, are ultimately responsible to each other for complying with these rules. This basic relation of mutual accountability exists even if the workers should choose to employ various hierarchical structures for the sake of efficiency. In the second conception, the firm's management determines the rules governing the workers' labor and it is to management alone that the workers are accountable. No basic relation of mutual accountability between the workers exists, even as the management might implement various team-building exercises and peer evaluation mechanisms. The managerial firm fails to be a genuinely cooperative scheme because its members lack standing, just in virtue of being part of

and reciprocity are accounted for only indirectly, as what is thought to be normally necessary to maximize the sum of social welfare."

⁵ The idea of a central coordinating authority here resembles Joel Feinberg's (1970) idea of a sovereign rights monopoly, but the two ideas are importantly distinct. Feinberg's concern is with individuals' standing to demand the performance of duties *owed to them*, since this standing is part of what explains the value of rights. But not every duty a person has standing to demand is a duty that is owed to or directed toward that person. Both standing and direction are important components of claim-rights, but they are distinct. I discuss the concept of directed duties in May (2015) and provide an account of the difference between standing and direction in a separate paper project.

the firm, to demand that the other members comply with its rules. It is instead a more or less complicated form of socially coordinated activity.

The second relational constraint, then, is that the principles of justice are justified only if they establish a fundamental horizontal relation of moral accountability between citizens. Utilitarianism does not satisfy this constraint since it does not explain why it is justifiable for citizens to have the moral standing to demand sacrifices of each other in pursuit of the goal of maximal desire satisfaction. The doctrine only claims society should be efficiently arranged so that these sacrifices are in fact made. The form relations of accountability take in a utilitarian society is, accordingly, entirely derivative of the principle of utility.

In contrast to utilitarianism, justice as fairness incorporates mutual accountability as a fundamental constraint on principles of justice. As an example, consider the criterion of reciprocity Rawls invokes in his defense of political liberalism. According to this criterion, proposed terms of association are fair only if their advocates may reasonably expect other citizens to reasonably accept them.⁶ On one articulation of the criterion, it is enough that the advocates *sincerely believe* other citizens can accept the terms “as free and equal, and not as dominated or manipulated, or under the pressure of an inferior political or social position” (2005, xlii). But on this articulation, the criterion does little to constrain the terms of association, since optimistic citizens could sincerely believe (even if, perhaps, incorrectly) that others might one day be persuaded to accept terms of association grounded in a comprehensive conception of the good. A sharper articulation of the criterion is that proposed terms are fair only if it is reasonable for their advocates to expect of other citizens that they accept the terms. Here, the reasonable expectation is normative rather than descriptive—it is not merely a sincere belief about what others might endorse, but a *justifiable demand* that they accept the terms in question. On this articulation, the criterion has much greater potential to constrain fair terms of association. Sincere

⁶ Rawls's liberal principle of legitimacy applies the criterion of reciprocity to the political structure: “the exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse” (2005, 137). I criticize the liberal principle as a conception of political legitimacy in May (2009).

beliefs do not themselves require moral justification—they are either true or false, supported by the evidence or not, and so on—but relations of mutual accountability do. Comprehensive conceptions of justice plausibly violate the criterion, given the burdens of judgment (Rawls 2005, 54–58), because it is not morally justifiable to demand of other citizens, on the political liberal view, that they endorse any particular comprehensive moral doctrine. Since citizens are not accountable to each other, as citizens, for endorsing a comprehensive doctrine, they cannot be accountable to each other for accepting any conception of justice that depends on such a doctrine.

II. Moral Demands in Cooperative Schemes

I have claimed that the cooperative conception of society that underpins justice as fairness includes the idea that participants in a fair scheme of cooperation are bound together by horizontal relations of mutual accountability: they all have standing to demand of each other that they comply with the rules governing the scheme. This provides the second premise for the cooperative interpretation of ideal theory:

(2) A scheme of genuine social cooperation exists only when individual participants are morally accountable to each other in that they each have standing to demand that the other participants of the scheme comply with its rules.

What does a horizontal relation of mutual accountability add to the basic idea that the scheme ought to be effectively governed by the rules in question? In this section, I advance an account of the nature of moral demands in cooperative schemes in response to this question.

To fix ideas, consider an example where one person plausibly lacks moral standing to demand that another perform her moral duty:

Strangers on a Train

Anne and Bruno are two strangers sitting next to each other on a train. Anne is talking on her phone, describing in great detail her plans to ruin her

friend Guy's romantic relationship by spreading deceitful rumors. Bruno follows her conversation for some time and forms the correct belief that Anne has a moral duty to not carry through with her malicious meddling.

I assume here that Bruno has no moral standing to demand that Anne abandon her plans, despite the truth of his belief. It would be out of place, in some sense, for him to turn to her and insist, "Anne, you must not ruin Guy's romantic relationship!" Her immoral meddling is, intuitively, none of his moral business.⁷ The reason for this lack of standing is not epistemic. Bruno might at first lack sufficient information about Anne's friendship with Guy to know whether any moral considerations could justify her action. But as the conversation progresses, it can become apparent the deceit is entirely unjustified. Nor is the problem that Bruno's assertion would be a separate action from whatever Anne chooses to do, and hence subject to moral evaluation itself. Bruno does not lack standing because his assertion would be a rude intrusion into her private conversation. Anne might invite him to offer his honest opinion, in which case his comment would not be rudely intrusive. Yet Anne would still not be morally accountable to him for her actions. Three proposition types can accordingly be distinguished:

- (i) X has a moral duty to ϕ .
- (ii) Y may justifiably assert that X has a moral duty to ϕ .
- (iii) Y has moral standing to demand that $X \phi$.

The first two propositions do not imply the third because they do not imply anything about the moral relationship between the two people. The possession of moral standing in (iii) adds something to the facts in (i) and (ii). But there are at least two accounts of what this additional element is: a normative power or a normative liability.⁸

⁷ Not everyone may share this intuition. Readers who believe that Bruno does have the moral standing to address this demand to Anne are invited to substitute their preferred example of a lack of standing.

⁸ Here I employ a broad concept of a normative power as the ability to change an agent's normative position by intentionally performing some action. An agent has a normative liability when some other agent has a normative power with respect to her.

The first account claims that *Y*'s moral standing in (iii) entails that he has the normative power to impose on *X* (or otherwise modify) a moral duty to ϕ by demanding that she ϕ . This *power theory* of moral demands explains Bruno's lack of standing in *Strangers* by claiming that his assertion ("Anne, you must not ruin Guy's romantic relationship!") would not change her moral situation. She would not become even more duty-bound to not spread deceitful rumors in virtue of Bruno's utterance. This is correct, as far as it goes, but the power theory is an inadequate theory of the standing to demand. The problem with the theory is that it places too much emphasis on the act of issuing a demand itself. A demand does sometimes involve a normative power to change its addressee's moral position. If an officer demands that a soldier retrieve some equipment, then the soldier *ipso facto* becomes duty-bound to do so. In this case, the demand is a command the officer has the authority to issue. But not all demands are commands. In many cases, demands forcefully communicate the addressee's existing duty, but do not add any force or substance to that duty. For instance, Anne has the moral duty to not punch Bruno, and he has the moral standing to demand that she comply with this duty. But no utterance of his to that effect would intensify (or otherwise change) her duty to not punch him.

The second account claims *Y*'s standing in (iii) entails that he possesses a normative liability with respect to *X* and at least one of his existing duties: if *X* were to violate her duty to ϕ , she would thereby suspend, *ceteris paribus*, *Y*'s duty to ψ .⁹ This *liability theory* of moral demands explains Bruno's lack of standing in *Strangers* by claiming Anne's plans to ruin Guy's relationship leave Bruno's moral position with regard to her unchanged in a particular way: he does not become permitted, even *ceteris paribus*, to sanction Anne for her conduct. In broad terms, a sanction is some form of negative treatment imposed on a person to enforce compliance with a duty, negative treatment that it is ordinarily impermissible to impose. Thus, *Y*'s duty to ψ is his ordinary or default duty to refrain from sanctioning *X*. Refer to the combined duties *Y* has with regard to *X* when his duty not to sanction her remains in force

⁹ I briefly defend this account of the standing to demand in May (2013).

as his *default duty of normal cooperation* with X.¹⁰ Bruno lacks standing to demand that Anne discontinue her meddling in Guy's relationship because her actions do not suspend his default duty of normal cooperation with her. He may think worse of her, decide not to be especially friendly toward her, or complain about her to his friends after the train journey. But these would not constitute departures from normal cooperation because they would not constitute sanctions it would ordinarily be impermissible for him to impose.

The key idea in the liability theory of moral demands is that when X is morally accountable to Y there is a link—a kind of tripwire—between one of X's moral duties and Y's default duty of normal cooperation: X's violation of her duty triggers, *ceteris paribus*, the suspension of Y's duty to not impose a sanction on her. Y's moral standing to demand that X ϕ consists in the fact that X's violation of her duty to ϕ would have this normative upshot for his moral position. A relation of mutual accountability exists when two agents are morally accountable to each other, although not necessarily in exactly the same way. Consider a revised example of the *Strangers* example, one where a relationship of moral accountability does plausibly exist:

Fellow Passengers on a Train

Anne and Bruno are fellow passengers sitting next to each other *in the quiet car* of a train. Anne is talking on her phone, describing in great detail her plans to ruin her friend Guy's romantic relationship by spreading deceitful rumors. Bruno follows her conversation for some time and forms the correct belief that Anne has a moral duty to end her phone call.

In this example, Anne has a duty to end her phone call because she is in the quiet car, where phone calls are explicitly forbidden. Talking on the phone in the quiet car is wrong, at least insofar as other passengers might be disturbed. As things go, however, it is not an especially egregious wrong. Anne's interference in Guy's romantic life is a far graver misdeed

¹⁰ I assume for simplicity that different duties can be combined into a single conjunctive duty.

than her lack of consideration for her fellow passengers. Nevertheless, I assume Bruno has the moral standing to demand that she end her conversation, even as he lacks the standing to demand that she end her meddling. Whereas he would ordinarily have some duty not to bother her with an insistent request—it's the quiet car, after all—her behavior changes what he may do. Should she refuse to comply with his request, he would become permitted to take further steps, such as summoning the conductor to reproach her. None of these departures from normal cooperation on Bruno's part would be permissible if Anne were not violating a rule against phone calls. He may not disturb her or cause a fuss just in virtue of the morally objectionable content of her conversation.

The difference between *Strangers* and *Fellow Passengers* can be explained in terms of the cooperative scheme in which Anne and Bruno participate. Passengers on a train find themselves together in a common venture: they must occupy the same physical space for the duration of a journey from one station to another. The passengers cooperate with one another by complying with certain explicit and implicit rules that respond to their shared interests as passengers. These *cooperative rules* constitute the moral terms of their temporary association on the train. Each passenger has, as a participant in the cooperative pursuit of a reasonably pleasant journey, standing to expect of the others that they comply with the terms of their association. One such term, in the quiet car at least, is the rule forbidding phone calls. This rule responds to the passengers' interests insofar as they are assumed to have, in virtue of selecting the quiet car, a strong preference for a peaceful trip and no strong interest in the freedom to converse as they like. But no moral rule forbidding malicious interference in non-passengers' romantic lives would be warranted as a response to their shared interests as passengers. With respect to such extraneous moral matters, the passengers are strangers. With respect to their behavior on the train, they are fellows.

The train examples are simplistic in at least two ways. First, not all moral rules governing train travel are a matter of cooperation with one's fellow passengers. In addition, train travel is a socially coordinated activity. Thus, there are typically rules governing how the passengers

are to behave that do not constitute terms of association between them. Rather, the passengers are only accountable to the train's staff for compliance with these *coordination rules*. For instance, each passenger is obligated to purchase a valid train ticket, just as every passenger in the quiet car is obligated to refrain from phone calls. But whereas the latter obligation stems from both types of rule, the former obligation plausibly stems from a coordination rule alone. On this supposition, Bruno does not have standing to demand that ticketless Anne seek out the conductor to pay her fare.¹¹ Second, members of the same society are participants in a great many cooperative schemes, with complicated boundaries and frequently nebulous terms of association. Even strangers occupying the same train car are not merely fellow passengers and hence do not only have standing to make moral demands of one another in that capacity. For instance, if Anne were to start beating her infant child, Bruno would have standing to demand that she stop. If she refuses to stop, he becomes permitted to restrain her. This is not primarily because he has an interest as a passenger in not hearing the child's wails. Rather, the relevant association is something deeper: as members of the same society, people may expect of each other that they not violently abuse infants. Bruno would have the same standing if he were walking past Anne and her child on the street.

Identifying the salient terms of association is often difficult—even if there is rough agreement on what moral duties people have, it can be hard to determine the constituency to which people are accountable for performing those duties. Disagreement about the contours of accountability relationships and the precise terms of normal cooperation generates much of the heat of moral conflict in common life. My present point is not to deny (still less settle) these complexities, but to defend the claim that a relationship of moral accountability with respect to certain duties is something additional to those duties and something that itself requires justification. In particular, moral accountability requires a normative tripwire between one agent's moral duties and another agent's default duties of normal cooperation with her, a tripwire that stands in need of

¹¹ I do not deny that there may be circumstances in which train passengers do have the moral standing to demand that they each purchase a ticket.

moral justification. This point would hold even if it were true—implausibly—that every person is always morally accountable to every other person for all their moral duties.¹²

The liability theory of moral demands provides the third premise for the cooperative interpretation of ideal theory:

- (3) An individual *Y* has moral standing to demand that another individual *X* comply with a duty only when *X*'s violation of this duty would trigger, *ceteris paribus*, the suspension of *Y*'s default duty of normal cooperation with her.

III. Strict Compliance in the Moral Theory of Cooperation

I have thus far presented three premises in the cooperative interpretation of ideal theory:

- (1) A just political society constitutes a grand scheme of social cooperation between free and equal persons for reciprocal advantage.
- (2) A scheme of genuine social cooperation exists only when individual participants are morally accountable to each other in that they each have standing to demand that the other participants of the scheme comply with its rules.
- (3) An individual *Y* has moral standing to demand that another individual *X* comply with a duty only when *X*'s violation of this duty would trigger, *ceteris paribus*, the suspension of *Y*'s default duty of normal cooperation with her.

In this section, I argue that premises (2) and (3) imply a further claim:

- (4) The moral terms of genuinely cooperative schemes depend on a strict compliance assumption.

¹² Linda Radzik (2011) raises a similar objection to the universal scope of Stephen Darwall's (2006) account of the second-personal standing to blame.

The argument for premise (4) turns on the interdependence of the duties in genuinely cooperative schemes. It comprises two parts. The first part establishes the logical need for a strict compliance assumption at some point in the justification of the scheme's rules. The second part establishes the justificatory priority of this assumption.

1. Strict Compliance and the Logic of Moral Demands

Premise (2) asserts that the participants of a cooperative scheme have standing to demand of each other that they comply with the scheme's rules. I also assume that if some individuals have standing to demand of each other that they comply with a social rule, then that rule is cooperative and constitutes one of the terms of their association. Other social rules may also govern the individuals' behavior, but without the required relation of mutual accountability, these would not constitute rules of a cooperative scheme that comprises the individuals as participants. Instead, they would be rules of some other scheme of social interaction.

The first part of the argument for premise (4) is quite intricate. It comprises six steps. The first step of the argument posits a link between the individuals' duties and the rules of their cooperative scheme:

- (i) A participant X of a cooperative scheme S has a duty to ϕ , *qua* participant in S , only if some cooperative rule R_1 in S imposes on X a duty to ϕ .¹³

Refer to the duties X has as a member of a cooperative scheme as her cooperative duties. X may have other duties that are not imposed by any of the rules of S and which are therefore not S -cooperative duties. These extraneous duties would be ones she has as a member of another scheme of interaction. They may be related to her S -cooperative duties in complicated ways—they could even be indirectly grounded in her identity as

¹³ I assume for simplicity that different social rules can be combined into a single conjunctive social rule. Nothing turns on whether there is only one identifiable norm in play or a number of different norms.

a participant in *S* and hence be duties she has only because of her *S*-cooperative duties. But if she is not accountable to the other members of *S* for compliance with these duties, then they are not duties she has *qua* participant in *S*.

Premise (3) incorporates the liability theory of the standing to demand. If some individual *Y* has the standing to demand that *X* comply with her *S*-cooperative duties, then he must have a default duty of normal cooperation with her—a default duty that is suspended, *ceteris paribus*, when *X* violates her *S*-cooperative duty to ϕ . But this default duty must itself be a cooperative duty, otherwise the rules of the scheme would be incomplete. This means there must be a cooperative rule of *S* that imposes on *Y* a default duty of normal cooperation with *X*. Thus, the second step of the argument posits a particular *S*-cooperative rule:

(ii) Some cooperative rule R_2 imposes on *Y* a default duty of normal cooperation with *X* on the assumption that she complies with her various *S*-cooperative duties.

The third step generalizes the second:

(iii) Some cooperative rule R_3 imposes on *Y* an array of default duties of normal cooperation with all other participants in *S* on the assumption that each of them complies with their *S*-cooperative duties.

Here R_3 entails R_2 as a subordinate rule, since *Y*'s default duties with regard to *X* follow from his array of default duties with regard to all the scheme's participants. In its simplest form, R_3 just universalizes R_2 —that is, it states that *Y*'s default duties of cooperation are the same with regard to all other participants of the scheme. In more complicated cases, R_3 imposes on *Y* different default duties with regard to different members of *S*. For instance, the default duties of normal cooperation adults have with regard to children presumably differ from those they have with regard to other adults. Whatever complexity is required here can be accommodated in the complexity of R_3 .

The fourth step narrows the scope of R_2 , since it identifies the default duties of normal cooperation *Y* has with regard to *X* on the assumption

that she complies not with all her *S*-cooperative duties, but only with the various default duties of normal cooperation she has with regard to the other participants of *S* in virtue of a corresponding rule of type R_3 .

(iv) Some cooperative rule R_4 imposes on *Y* a default duty of normal cooperation with *X* on the assumption that she complies with her own array of default duties of normal cooperation with all other participants of *S*.

The fifth step generalizes the fourth, just as the third step generalizes the second:

(v) Some cooperative rule R_5 imposes on *Y* an array of default duties of normal cooperation with all other participants in *S* on the assumption that each of them complies with their default duties of normal cooperation with all other participants of *S*.

The final step generalizes once again:

(vi) Some cooperative rule R_6 imposes on all participants of *S* an array of default duties of normal cooperation with all other participants of *S* on the assumption that each of them complies with their own array of default duties of normal cooperation with all other participants of *S*.

The generalization in the sixth step allows for similar complexity to that in the third. Just as the default duties adults have with regard to children may differ from those they have with regard to other adults, the default duties of children may differ from those of adults. The necessary complexity can be accommodated in R_6 . Note, though, that however complex the rule may be, and whatever its particular content, it presupposes that all the participants of the cooperative scheme comply with its rules. This is because if all members comply with their default duties, there is no violation of a rule that could trigger the permissibility of sanctions or the applicability of rules governing the imposition of sanctions. In this case, the only applicable rules governing the participants are R_6 and the various subordinate rules it entails. This means that whatever a full

specification of the rules of *S* may ultimately contain, it contains at least one rule that applies to the participants when they all comply with the applicable rules of the scheme. The justification of R_6 must accordingly proceed on the basis of a strict compliance assumption. Refer to a rule of type R_6 as a *strict compliance rule*.

Note that a strict compliance rule is worked out on the assumption that all the participants of a scheme of cooperation comply with their duties. It does not follow that the rule ceases to have any normative purchase as soon as one participant violates a duty. A rule that ceased to have normative purchase as soon as it was violated would be unworkably fragile. Instead, strict compliance rules are more or less robust, depending on the scheme in question, insofar as they generate duties that hold across a more or less extensive range of circumstances, from the core case of strict compliance to various cases of partial compliance. A fully moral theory of the scheme in question will include auxiliary rules that specify how the participants' duties are affected the more non-compliance there is.

To illustrate the idea of a strict compliance rule, consider the terms of association in *Fellow Passengers*. The strict compliance rule is the rule that Anne, Bruno, and the other passengers are bound to follow on the assumption that all the occupants of the quiet car follow that same rule. It might have something like the following content:

The Quiet Car Rule

Sit quietly in your seat for the duration of your journey. Do not disturb other passengers on leaving your seat or otherwise physically annoy or pester them. Do not talk on the phone or to yourself, or otherwise make any noise that might disturb other passengers. If you wish to listen to music, use headphones and keep the music at a volume other passengers cannot hear. Do not try to engage other passengers in conversation. If for some pressing reason you must communicate with someone else on board the train, do so very briefly and in a low voice.

The quiet car rule is a strict compliance rule since it supposes that the passengers all comply with it. It does not say how Bruno should respond to Anne's refusal to end her phone call, at least after a whispered request

to her proves ineffective. Her continued violation of the quiet car rule leaves him in a somewhat awkward predicament. He does not become permitted to do just anything to get her to comply with the rule—presumably, he may not grab her phone and throw it out the window. But nor is he necessarily required to sit silently and endure her lack of consideration. Thus, a full specification of the cooperative rules of the quiet car would include some further rules providing guidance to the passengers in cases where some of their number fail to abide by the strict compliance rule.

2. The Priority of Strict Compliance

As the argument stands, every scheme of genuine social cooperation requires a strict compliance rule of type R_6 . It does not follow, however, that this strict compliance rule is anything other than an artificial way to articulate the logical implications of a relationship of mutual accountability in a cooperative scheme. What real work does the strict compliance rule do in the justification of the rules of S as a whole? There are two problems here. First, the strict compliance rule must have some sort of priority in the moral justification of the overall rules of the scheme. Without this priority, the strict compliance puzzle cannot be solved. Ideal theory could be a philosophically defensible endeavor, but there would be no reason to accept Rawls's claim that ideal principles of justice provide the necessary foundation for moral norms governing responses to injustice. Refer to this as the *priority problem*. Second, the strict compliance rule must be justified on the basis that the participants' overall array of default duties of normal cooperation constitute a fair and reciprocal scheme of cooperation in itself. If the values of fairness and reciprocity only governed the scheme's rules once the strict compliance assumption was suspended, then there would be no important substantive content in the role played by ideal theory. Refer to this as the *fairness problem*.

The solution to the priority problem draws on the idea that default duties of normal cooperation always include a duty to refrain from sanctioning a person for some violation. A sanction is not just any sort

of negative treatment—it is not always ordinarily impermissible to treat another participant in a cooperative scheme in a negative manner. Rather, sanctions have a particular rationale: their imposition is warranted only in response to some (threatened) violation of a duty. This means all rules governing what sanctions may be imposed, who may impose them, what must be done once they are imposed, and so forth are ultimately secondary, either directly or indirectly, to the duties participants have when the question of sanctions does not arise. But these are the duties the participants have when no one violates a duty; that is, when they all comply with the rules of the scheme. And the duties they have when all comply with the rules of the scheme are the default duties of normal cooperation stemming from the strict compliance rule. Thus, a strict compliance rule is the core element of any system of rules for a cooperative scheme.

The solution to the fairness problem turns on a particular kind of interest at stake in the justification of any fair scheme of cooperation. A cooperative scheme is fair only when it properly promotes or satisfies the various relevant interests of the participants. Not every individual interest is relevant, to be sure, but the question of fairness does not arise without some reference to the participants' interests. Some of the relevant interests are typically non-normative, in the sense that their fulfillment makes a person's life go better independently of any normative state she may be in. For instance, people have interests in being healthy and happy, but health and happiness are not in themselves normative states. But individuals also have normative interests; that is, interests in being in a certain normative state (Owens 2012). For instance, people have an interest in possessing the normative power to permit others to enter their homes. Another normative interest helps resolve the fairness problem. This is the normative interest a participant in a cooperative scheme has in being a member in good standing of the scheme. For an individual to be a member in good standing in a scheme is for all participants' default duties of normal cooperation with her to be in force, and for this fact to be generally recognized by the participants. The interest in good standing is distinct from any interest the person has in not being sanctioned by the other participants for some (perceived) violation of a duty—the value of good standing to her is not exhausted by how it

makes such negative treatment less likely. Indeed, it is very often the threat to this normative interest in itself that motivates compliance with the scheme's rules—people are frequently more perturbed by the mere prospect of being in moral trouble with their peers than of experiencing any negative treatment that may arise in consequence (May 2013, 559).

The normative interest in good standing is a natural element of the Rawlsian cooperative conception of society.¹⁴ Rawls identifies two moral powers: the capacity for a conception of the good and the capacity for a sense of justice (1971/1999, 41). The sense of justice—the effective desire to comply with the rules of a just scheme—corresponds to the interest in good standing. This is because when citizens recognize a person is effectively motivated to act in accordance with just rules, they recognize their default duties of normal cooperation with her remain in force. The function of a sense of justice in justice as fairness is not simply to explain how citizens might freely comply with the institutions of a well-ordered society. In addition, the principles of justice are themselves justified, in part, by how well they allow for the proper development and realization of a sense of justice. Other things being equal, a conception of justice that stymied the sense of justice would be deficient. One reason for this is that the parties in the original position are motivated to ensure the social bases of self-respect, but cannot do so without providing for the interest in good standing. This is because a person's self-respect depends, in part, on the respect she receives from others, and to respect another person is to regard her “as a moral being with a sense of justice and a conception of the good” (1971/1999, 296). People are often indignant when they are unjustifiably accused of violating just terms of association primarily because such an unfair accusation is disrespectful, not because it may lead to some sanction in consequence.

More generally, the interest in good standing is always a relevant interest in the justification of the rules of a fair scheme of cooperation. A relationship of mutual respect is a corollary of a relationship of mutual accountability—the participants respect each other insofar as they are

¹⁴ In addition, Rawls (1999b) implicitly appeals to the interest in his defense of the claim that decent hierarchical peoples should be recognized as full members in good standing of the international Society of Peoples.

prepared to hold each other accountable for following the rules of the scheme. Thus, the fairness of a cooperative scheme requires an assessment of its fairness when each participant's interest in good standing is satisfied. But this occurs only when everyone's default duties of normal cooperation remain in force; that is, in the circumstances governed by the strict compliance rule. This means the strict compliance rule must be justified on the basis that the default duties of normal cooperation themselves constitute a fair and reciprocal cooperative system. Unless this constraint is satisfied, the scheme as a whole could only fairly balance the various other interests at stake at the cost of the interest in good standing of at least some of the participants. Fairness and mutual respect would be at loggerheads.

The fairness of a cooperative scheme's strict compliance rule has an important consequence. Violations of the rules of the scheme can be correctly criticized as unfair. For instance, in *Fellow Passengers*, Anne unfairly takes advantage of the quiet provided by the other passengers by violating the quiet car rule. But violations of the rules of the scheme must be unfair in virtue of departing from some standard of fairness. This standard of fairness cannot itself be a standard that assumes violations of the scheme's rules, since then these violations would not constitute departures from fairness. Rather, they would be preconditions of the scheme's overall fairness. Thus, criticism of Anne's behavior as unfair refers back, ultimately, to the fairness of the quiet car rule. More generally, then, the strict compliance rule possesses justificatory priority with respect to the justification of the overall scheme. The moral terms of a cooperative scheme therefore depend on the justification of its strict compliance rule. Since this rule depends on a strict compliance assumption, premise (4) follows:

- (4) The moral terms of genuinely cooperative schemes depend on a strict compliance assumption.

IV. The Well-Ordered Society

I have argued that moral theories of cooperation require strict compliance assumptions because cooperative schemes involve relations of

mutual accountability between the participants. Premise (1) and premise (4) together imply that principles of justice should be justified as an exercise in ideal theory:

(1) A just political society constitutes a grand scheme of social cooperation between free and equal persons for reciprocal advantage.

(4) The moral terms of genuinely cooperative schemes depend on a strict compliance assumption.

Therefore:

(5) The moral terms of a just political society—its principles of justice—depend on a strict compliance assumption.

In this section, I illustrate the cooperative interpretation of the ideal theory of justice by discussing two features of Rawls's statement of justice as fairness: two roles played by the principles of justice of a well-ordered society and the connection between ideal theory and Rawls's restriction of the principles of justice to the basic structure of society.

1. Two Roles for Principles of Justice

In section 1 of *A Theory of Justice*, Rawls discusses the role of the principles of justice as follows:

Let us assume, to fix ideas, that society is a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them. Suppose further that these rules specify a system of cooperation designed to advance the good of those taking part in it. Then, although society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests.... A set of principles is required for choosing among the various social arrangements which determine [the] division of advantages and for underwriting an agreement on the proper distributive shares. These principles are the principles of social justice: they provide

a way of assigning rights and duties in the basic institutions of society and they define the appropriate distribution of the benefits and burdens of social cooperation. (1971/1999, 4)

Rawls's statement of role of the principles of justice includes two elements. The first and most obvious role is *distributive*. The principles of justice set out the fundamental moral criteria for the distribution of benefits and burdens between members of society. This is a role justice as fairness shares with utilitarianism, since the latter also includes an account of how benefits and burdens ought to be distributed. More generally, every theory of justice includes principles that play this distributive role.

The second, *regulative* role is less obvious. Principles of justice determine the rules of association that the members of society "in their relations to one another recognize... as binding." The binding force of the principles of justice, that is, does not lie simply in the fact that they govern the distribution of benefits and burdens. In addition, the principles are binding in the moral relations between citizens. I interpret this to mean that in a just society, the citizens are bound together by a certain relation of mutual accountability: citizens are accountable to each other for complying with the institutional rules implied by the principles of justice. This is not a role for the principles that justice as fairness shares with utilitarianism. In the utilitarian approach, the principles of justice may simply bind a central coordinating authority in its imposition of various social arrangements. Whether or not the members of a utilitarian society should recognize the principle of utility as binding in their moral relations with each other is a quite separate matter.

The two roles can be expressed in terms of two questions. In their distributive role, the principles answer the question, "what, as members of society, do we owe to each other?" In their regulative role, the principles answer the question, "what, as members of society, may we demand of each other?" A well-ordered society constitutes a grand scheme of cooperation, then, when it is governed by principles that provide a unified answer to these two questions. On this conception of justice, what we owe each other is the same as what we can demand of each other. My claim here is that it is in virtue of the regulative role of the

principles of justice that they must be worked out, in justice as fairness, on the basis of a strict compliance assumption.

2. Ideal Theory and the Basic Structure Restriction

In section 2 of *A Theory of Justice*, Rawls imposes two limitations on the subject of justice. First, the principles of justice are to be developed as an exercise in ideal theory. Second, they apply to the basic structure of society; that is, “the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation” (1971/1999, 6). As an exegetical matter, the combination of these two restrictions in a single section should seem slightly curious. On the telic approach to ideal theory, there is no obvious reason why principles of justice should take the basic structure of society as their subject. If the well-ordered society is the perfectly just society political decision makers should aim at, then they should aim at it whatever the scope of its principles. Consider G. A. Cohen’s (2008) critique of the basic structure restriction. On Cohen’s view, the individuals in a just society would make certain types of decisions that are not required by the society’s rules. For instance, they would freely choose to forgo incentives that shift society away from an egalitarian distribution. For Cohen, principles of justice identify a state of affairs in which each person receives what she is owed. The basic institutions of society are important subjects of justice, as Cohen recognizes, but only because they play an important role in bringing about the appropriate distribution. Since an egalitarian ethos can play this role too, there is no good reason to single out the basic structure as special. Both institutions and free choices are important as strategic routes to the realization of the ultimate telos.

The cooperative interpretation of ideal theory, in contrast, provides at least some rationale for Rawls’s basic structure restriction. A preliminary argument to this effect can be briefly sketched as follows. The principles of justice are the terms of association of the well-ordered society—in their regulative role, they define what citizens may demand of each other as citizens. But the duties citizens have standing to demand of each other

are cooperative duties, and these duties must stem from the cooperative rules of the society. Thus, there are no requirements of justice that do not stem from the cooperative rules of a well-ordered society. The basic structure of society comprises those institutions that establish the fundamental rules of society (Rawls 1971/1999, 47). All other social institutions establish social rules only in accordance with the rules of the basic structure. Whatever citizens may demand of each other as a matter of fundamental justice, they may demand of each other only in accordance with the institutional rules of the basic structure.¹⁵ Thus, if the principles of justice are to have a regulative role, as required by the idea of relations of mutual accountability, they should apply in the first instance to the basic structure. Principles of justice for other institutions, associations, and practices are subordinate to the principles governing the basic structure. And if the principles of justice for the basic structure are the fundamental terms of association, establishing the fundamental horizontal relation of mutual accountability, then they must be justified on the basis of a strict compliance assumption.

V. An Anti-Utopian Objection to Ideal Theory

In justice as fairness, the idea of a well-ordered society has two main features. First, it presents a substantive ideal of fair and reciprocal social cooperation between free and equal persons for mutual advantage. In particular, it provides an account of a society in which citizens have the moral standing to hold each other accountable for acting in accordance with the principles of justice. Second, the well-ordered society is a realistic utopia: it stands as a political telos that should orient decision making in the actual world. Once the end of a well-ordered society has been specified, the appropriate means to achieve it can be determined. On the cooperative interpretation of the ideal theory of justice, the strict compliance assumption is grounded in the first of these features. All moral theories of cooperation have a strict compliance rule at their core,

¹⁵ For a related response to Cohen's objection to the basic structure restriction, see Williams (1998).

and, in the Rawlsian approach, a theory of justice is just a theory of social cooperation writ large. On the telic interpretation, in contrast, the strict compliance assumption is grounded in the second feature. If the well-ordered society is to properly fulfill its function as a realistic utopia, it must be assumed to be perfectly just. The substantive nature of the well-ordered society—in particular, the idea of mutual accountability it incorporates—forms no part of the rationale for ideal theory. The key difference between the cooperative and telic approaches therefore turns on which feature of the well-ordered society grounds the strict compliance assumption and which feature is, in consequence, inessential. This difference can be expressed in terms of two competing claims:

- (i) Since the telic function of a realistic utopia grounds the strict compliance assumption in justice as fairness, the substantive nature of a well-ordered society is inessential to ideal theory.
- (ii) Since the substantive nature of a well-ordered society grounds the strict compliance assumption in justice as fairness, the telic function of a realistic utopia is inessential to ideal theory.

In this section, I discuss an anti-utopian objection to ideal theory. Because this objection assumes the truth of the first claim, it does not threaten the cooperative interpretation. In the final section, I discuss the second claim and defend the possibility of non-utopian ideal theory.

Amartya Sen (2006) presents an objection to Rawls's ideal theoretic approach to justice. Sen claims political decisions are made as choices between some set of available options: *a*, *b*, *c*, ... etc. The decision makers should compare these options and select the best one. To do this, they do not need to identify some perfect state of affairs, *z*, that is not presently available to them. Since that option is unavailable to them, it is irrelevant to the assessment of the merits of *a*, *b*, *c*, etc. In Sen's analogy, it is unnecessary, and quite pointless, to identify the tallest mountain in the world before we determine the relative heights of Kanchenjunga and Mont Blanc (2006, 22). For Sen, justice is a matter of comparative assessments: decision makers should make improvements in the actual world by moving from one state of affairs to a better state of affairs. A conception of a perfectly just society is useless in that respect.

Note that Sen's objection to ideal theory does not depend on any substantive disagreement with Rawls's principles of justice. A comparative approach could rank the available options on the basis of three (lexically ordered) criteria: which option best guarantees equal basic liberties?; which option best promotes equality of fair opportunity?; and which option most improves the life prospects of the worst-off group? The disagreement between Rawls and Sen about ideal theory appears to be generic, in the sense that it does not turn on their substantive disagreements about the appropriate content of principles of justice. Sen's objection therefore assumes that the substantive nature of a well-ordered society forms no part of Rawls's rationale for ideal theory. Rather, he objects to the idea that the theory of justice requires the identification of an ideal Paradise Island.

The appearance of a purely generic debate between proponents and opponents of ideal theory is bolstered by John Simmons's (2010) defense of Rawls against Sen's objection. Simmons rejects the principles of justice as fairness, but not Rawls's distinction between ideal and non-ideal theory: "it is not Rawls's employment of the distinction itself... that should be the target of those who find unconvincing Rawls's (or Rawlsian) political philosophy" (2010, 6). Simmons offers two responses to Sen. The first response claims that ideal theory is necessary because political decision makers require some account of the perfectly just society *z* to orient their selection between *a*, *b*, *c*, etc. For instance, *a* may appear to be more just on its own merits than *b*, but it does not follow that justice requires that *a* be preferred to *b*. This is because *a* may be a cul-de-sac. If society were to go down that road, it could get stuck in a state of significant residual injustice. In contrast, *b*, for all its immediate faults, might put society on a much surer road to a fuller realization of justice. Decision makers need to keep their eyes on the prize, and to do that, they need a vision of a perfectly just society. Simmons's second response is that the strict compliance assumption is necessary so that different conceptions of justice can be compared in themselves—complications arising from non-compliance muddy the philosophical waters.

Since my present goal is not to criticize the telic interpretation of ideal theory, I do not argue here that Simmons's responses to Sen fail. But

there are at least some reasons for skepticism. With respect to the first response, a comparative approach can easily recognize short-term gains can come at the cost of long-term goals. Only a very short-sighted comparative analysis would lead into moral *culs-de-sac*. Moreover, even if some distant lodestar is needed to orient current political decision making, there is no straightforward reason why that lodestar must be a perfectly just society, governed by principles developed on the basis of a strict compliance assumption. It would plausibly be enough for it to be a society with institutions that should be established as permanent fixtures, taking the frailties of human virtue and the expected vicissitudes of compliance into account. The reasonably just society could act as a *telos* for political decision making just as well as a perfectly just society. With respect to the second response, an estimate of expected levels of compliance with the institutions required by proposed principles of justice may well be integral to an evaluation of those principles. If the conception requires institutional arrangements that are very likely to be flouted, then the conception may well be, to that extent, deficient as a moral guide to action. On this view, there is little practical point in being guided by a vision of a perfect society with impracticable institutions.¹⁶

However well Sen's objections may tell against ideal theory on the *telic* interpretation, they do not threaten the cooperative alternative. The key point here is that the cooperative interpretation provides an account of ideal theory that is grounded in the substantive content of justice as fairness. Even if there is no generic need for political decision making to be guided by some or other vision of a perfectly just realistic utopia, the strict compliance assumption can be defended as part of the idea of genuine social cooperation. Since principles of justice have a regulative role, justice is not simply a matter of determining which of the available options is best according to some distributive metric. A just society is one in which citizens have the moral standing to hold each other accountable for acting as just institutions require. But if the principles of justice for such a scheme of cooperation must be developed on the basis of a strict compliance assumption, as I have argued, then an account of the

¹⁶ For an important argument against the claim that unrealistic theories of justice are deficient, see Estlund (2019).

perfectly just society follows in short order. Once the well-ordered society is identified as an ideal scheme of fair and reciprocal social cooperation, it should then act as a lodestar for political decision making. This defense of ideal theory therefore rejects the first of the two competing claims outlined above—the idea of a realistic utopia may include a strict compliance assumption, but only in virtue of a substantive appeal to the horizontal relations of mutual accountability in a genuinely cooperative well-ordered society.

VI. Non-Utopian Ideal Theory

Rawls's well-ordered society is utopian not just because it serves to orient political decision making. In addition, it is utopian because it is valuable as an end, rather than as a means to some other end. Since a just society is non-instrumentally valuable, perfectly just institutional arrangements should be permanent fixtures, with whatever adaptations to changing circumstances are required over time. On the telic interpretation of ideal theory, the only rationale for the strict compliance assumption is to identify the nature of a realistic utopia that stands as the ultimate end of political decision making. There is no place for the assumption except within this destination-setting business. But not every scheme of social interaction is valuable as an end. Some schemes are valuable only as means to some other end and may accordingly cease to exist once that end has been achieved. If these schemes are to be genuinely cooperative, however, they should still be governed by fair terms and horizontal relations of mutual accountability. On the cooperative interpretation of ideal theory, the moral terms of every genuine scheme of social cooperation must be grounded in a strict compliance assumption, however temporary or transitional the scheme may be. This supports the second of the two competing claims introduced in the previous section: since the substantive nature of a well-ordered society—in particular its horizontal relations of moral accountability—grounds the strict compliance assumption in justice as fairness, the idea of a realistic utopia is inessential to ideal theory. Ideal theory is not essentially utopian.

In *Fellow Passengers*, the quiet car rule is a perfectly ordinary example of an ideal theory. It specifies how passengers in the quiet car are to behave on the assumption that all behave as they should. In itself, this rule says nothing about how the passengers should behave if some of their number violate the rule. Thus, a complete moral theory of cooperation would include norms governing how the passengers should respond to non-compliance. The quiet car rule is presumably robust in the sense that it is still a relevant norm if one or two passengers make a noise. In this case, the other passengers have standing to demand that they comply with the rule and return the compartment to a state of reasonable silence. Here, a non-ideal moral theory of conduct in the quiet car is needed to guide the passengers in how they respond to their inconsiderate peers. At some point, however, extensive non-compliance with the quiet car rule will cause the scheme to collapse. If no one else in the quiet car appears to be paying any attention to the rule, it becomes defunct. In this case, Bruno loses the moral standing to demand that Anne end her phone call. The cooperative terms governing their temporary association are now presumably just the same as in all other compartments.

Consider a second everyday example:

Lunch Counter

Charlize is one of a large number of workers who all wish to order lunch at a window counter staffed by a single person. Each worker has an interest in placing their order as quickly and as smoothly as possible. Charlize is unsure about which moral principles govern how she might go about getting her lunch, taking the interests of the other workers into account.

The obvious response to Charlize's moral predicament is that the workers should cooperate with each other so that they may all get their lunches in a fair and expeditious manner. They ought therefore to constitute a cooperative scheme, governed by a fair rule, and Charlize ought to play her part by complying with this rule. For instance, the workers could cooperate with each other by adhering to the following simple rule:

The Lunch Queue Rule

Form a queue by standing in a single file behind the last person to arrive before you and allow the people ahead of you in the queue to be served before you.

The cooperative moral solution to the problem of *Lunch Counter* is an ideal theory, since it specifies what Charlize should do on the assumption that the other workers also comply with the rule. The lunch queue rule is robust in much the same way as the quiet car rule. It is still a relevant norm if some of the workers start pushing in line or circumventing it entirely. In this case, the other workers have standing to demand that they get to the back of the queue like everyone else and wait their turn. At some point, though, widespread non-compliance will cause the scheme to collapse, and Charlize is left with only non-cooperative moral solutions to her predicament. For instance, a non-ideal moral theory might allow her to squeeze her way through the crowd, as long as she did not elbow or shove any of the other workers out of her way.

Neither the quiet car rule nor the lunch queue rule is utopian in any sense. Both should serve to orient the decision making of Anne, Bruno, Charlize, and their fellows, but not as any kind of ultimate end. The association of the passengers on the train and the workers in the cafeteria is temporary. Much the same point holds in the political domain. Not every political scheme of cooperation should be established as a permanent fixture of society. Some have value only as temporary arrangements. Moreover, some temporary political arrangements have value only as transitional mechanisms to a better future. Even as transitional arrangements, however, these schemes raise important questions of fairness and reciprocity. In addition, there can be good reason for them to include horizontal relations of mutual accountability. If so, the moral terms for such transitional schemes should be worked out on the basis of a strict compliance assumption.

To illustrate the possibility of a non-utopian ideal political theory, consider the predicament that a society might find itself in the day a reviled dictator is forced from office. Utopian principles of political legitimacy define the nature of the (presumably democratic) political

system that should be established as an enduring set of institutions. But these principles do not immediately dictate how the transition to democracy ought to take place. Democratic legitimacy requires that citizens be allowed to select their representatives in free and fair elections, but citizens cannot just vote in an election the day after the dictator flees. So some morally justifiable transitional mechanism must be implemented, and this mechanism cannot be full democratic government. But the morally relevant considerations are not exhausted by questions of efficiency in transitioning to democracy and side constraints imposed by basic human rights. There is also an important question of fairness in how the transition should be effected. It would be unfair, *ceteris paribus*, for one political party to seize power in a coup and unilaterally impose its preferred conception of democratic arrangements, even if this conception were the best. A coup would circumvent appropriate deliberation and negotiation about various important issues: the structure of the electoral system to be used in a democratic election, the authority of the newly elected assembly to alter the constitution, the implementation of much-needed legislation in the interim, and so forth. Thus, the political parties find themselves in a predicament analogous to that faced by Charlize and her fellow workers.

Democratic Transition

The Party of Justice and Democracy (PJD) is one of several political parties that all wish to see a transition from dictatorship to some form of democratic government. Each party has an interest in effecting a transition to its preferred form of democracy as quickly and as smoothly as possible. The PJD is unsure about which moral principles govern how it might go about this transition, taking the legitimate interests of the other parties and their constituencies into account.

As in *Lunch Counter*, the obvious solution to the PJD's predicament is that the parties should cooperate with each other, on fair terms, in making decisions about the various preconditions of a successful transition. For instance, suppose they should cooperate with each other in accordance with the following rule:

The Transitional Government Rule

Allow all parties to be represented in a supreme decision-making forum. Accept that all decisions about the forthcoming election are to be made by a super-majority of parties in the forum, and that no party may enforce any decision not made by the forum. Accept that an executive committee of non-aligned individuals selected by the forum should govern the country prior to the election. Do not threaten campaigns of violence or intimidation if the forum does not reach certain decisions.

A rule of this sort constitutes a moral theory developed on the assumption that all parties behave as they should. This means it constitutes an ideal theory of transitional government. The transitional government rule is robust in the sense that if some recalcitrant party threatens to violate it, the other parties have the moral standing to demand that it fall back in line. If extensive non-compliance with the rule causes the forum to collapse, however, the PJD must seek out a quite different non-ideal moral solution to its predicament. At this point, it may even become morally permissible for the party to seize power in a bloodless coup, if there is no feasible way to establish a cooperative scheme of political decision making prior to a democratic election. The transitional government rule is accordingly a political ideal theory that has some moral purchase in transitional circumstances. But it very clearly cannot be a utopian ideal theory, since no conception of transitional government can serve as an ultimate end to orient decision making.

Conclusion

Rawls defines ideal theory in terms of a strict compliance assumption and claims that it takes priority over the non-ideal theory governing responses to injustice. The standard interpretation of Rawlsian ideal theory is telic: it supposes that the function of the strict compliance assumption is to define a conception of a perfect society—a realistic utopia—that should serve as the ultimate telos of political decision making. I have defended a different account of the strict compliance assumption. On the cooperative interpretation of ideal theory, the

assumption is grounded in the fundamental idea of society as a fair scheme of cooperation between free and equal citizen. The well-ordered society provides a vision of society as a grand scheme of cooperation in which citizens possess the moral standing to press demands of justice against one another. The well-ordered society may also act as a realistic utopia, but its moral justification depends on a strict compliance assumption in just the same way as the justification of all other genuine cooperative schemes—whether these be permanent or temporary, ultimate or transitional, grand or petty, paradisaic or mundane.¹⁷

References

- Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.
- Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
- Estlund, David. 2019. *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton, NJ: Princeton University Press.
- Feinberg, Joel. 1970. "The Nature and Value of Rights." *Journal of Value Inquiry* 4 (4): 243–57.
- Laden, Anthony Simon. 2013. "Ideals of Justice: Goals vs. Constraints." *Critical Review of International Social and Political Philosophy* 16 (2): 205–19.
- May, Simon Căbulea. 2009. "Religious Democracy and the Liberal Principle of Legitimacy." *Philosophy and Public Affairs* 37 (2): 136–70.
- May, Simon Căbulea. 2013. "What We May Demand of Each Other." *Analysis* 73 (3): 554–63.

¹⁷ Earlier versions of this chapter were presented at the 2018 Bled Philosophical Conference, the 2019 New Orleans Invitational Seminar in Ethics, the 2019 Philosophy Desert Workshop, and the 2019 Workshop for Oxford Studies in Political Philosophy. I am grateful to the participants for their responses, and especially to Kwame Anthony Appiah, Rowan Cruft, David Estlund, Sally Haslanger, Blain Neufeld, Jon Quong, John Schwenkler, Amy Sepinwall, Olufemi Taiwo, Marshall Thompson, Bas van der Vossen, the editors of this volume, and two anonymous referees for their comments. I am also grateful to the graduate students in my seminar on ideal theory at Florida State University in the fall semester of 2018 for helpful discussions of the subject. I regret not being able to respond to all of the many criticisms that have been raised against earlier drafts.

- May, Simon Căbulea. 2015. "Directed Duties." *Philosophy Compass* 10 (8): 523–32.
- Neufeld, Blain. 2017. "Why Public Reasoning Involves Ideal Theorizing." In *Political Utopias: Contemporary Debates*, edited by Kevin Vallier and Michael Weber, 73–93. Oxford: Oxford University Press.
- Owens, David. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.
- Radzik, Linda. 2011. "On Minding Your Own Business: Differentiating Accountability Relations within the Moral Community." *Social Theory and Practice* 37 (4): 574–98.
- Rawls, John. 1958. "Justice as Fairness." *Philosophical Review* 67 (2): 164–94.
- Rawls, John. 1971/1999. *A Theory of Justice*. Revised edition. Cambridge, MA: Harvard University Press.
- Rawls, John. 1999a. *Collected Papers*. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press.
- Rawls, John. 1999b. *The Law of Peoples*. Cambridge, MA: Harvard University Press.
- Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.
- Rawls, John. 2005. *Political Liberalism*. Expanded edition. New York: Columbia University Press.
- Robeyns, Ingrid. 2008. "Ideal Theory in Theory and Practice." *Social Theory and Practice* 34 (3): 341–62.
- Sen, Amartya. 2006. "What Do We Want from a Theory of Justice?" *Journal of Philosophy* 103 (5): 215–38.
- Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy and Public Affairs* 38 (1): 5–36.
- Stemplowska, Zofia, and Adam Swift. 2014. "Rawls on Ideal and Nonideal Theory." In *A Companion to Rawls*, edited by Jon Mandle and David Reidy, 112–27. Hoboken, NJ: Wiley.
- Valentini, Laura. 2012. "Ideal vs. Non-Ideal Theory: A Conceptual Map." *Philosophy Compass* 7 (9): 654–64.
- Williams, Andrew. 1998. "Incentives, Inequality, and Publicity." *Philosophy and Public Affairs* 27 (3): 225–47.

Index

For the benefit of digital users, table entries that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.

- ableism 37
- accountability 234–5
 - moral 235, 239, 241–2, 258
 - mutual 234–6, 243, 249–55, 257–8, 260
- Acemoglu, D. 13
- Ackerman, Bruce 186
- Adams, Raven 157n.22
- Afnan, Max 157n.22
- Agarwal, Bina 32n.14
- Allen, Danielle 24–5, 39
- allocation branch 177–8
- Althusser, Louis 25–7, 41
- anarchic 213–14
- Anderson, Elizabeth 3–4, 27–8, 49n.25, 51–2, 56
- Anti-Vaxxers 52–3
- Antony, Louise 58n.29
- Anzaldúa, Gloria 47
- Arendt, Hannah 4, 100n.1
- Aristotle 194–5, 194n.1
- Arneson, Richard J. 16n.7, 194
- Arnold, Matthew 5
- Arpaly, Nomy 202–3
- Arrow, Kenneth J. 169–70
- Arrow-Debreu general equilibrium model 171–2
- Ásta 58n.29
- Atiq, Emad 58n.29
- authoritarian 100
- authority 3, 27, 42–3, 102–3, 110, 117, 146, 173, 180n.10, 184, 188–9, 194–6, 201–4, 206–7, 207nn.15–16, 208–15, 211n.23, 212nn.25–6, 216–17, 217n.31, 218–19, 220n.35, 221–3, 234, 234n.5, 238, 252, 260–1
- claim(s) to 195–6, 203–4, 206–7, 212n.25, 213–14, 217–18, 221–2
- democratic 221
- equal 194–5, 219, 222
- legitimate 145–7, 204n.12
- normative 2
- political 211n.23
- share(d) 195–6, 206, 209, 211–12, 211n.23, 212–13, 219, 221, 221n.37
- autocracy 208–9, 212–14
- Autor, David H. 13–14
- Badano, Gabriele 167n.4
- Bajaj, Sameer 121n.46
- Balkin, Jack 33–4
- Banerjee, Abhijit 168–9
- Barnes, Elizabeth 37
- Bartky, Sandra Lee 26, 43–4
- Bauer, Nancy 58n.29
- Bazargan-Forward, Saba 67, 77–8, 80–1, 83
- Beerbohm, Eric 223n.39
- Beitz, Charles R. 219–20
- Berger, Mark 223n.39
- Bessen, J. E. 12
- Bildung 6
- Blake, William 5
- Boettcher, J. 154n.18
- Boettke, Peter J. 169–70
- Bonotti, Matteo 167n.4
- Boulcault, Marion 58n.29
- Brady, S. T. 49n.27

- Brandt, R. 114n.31
 Brettschneider, Corey 223n.38
 Brison, Susan 58n.29
 Broad Scope View 67, 83, 94
 Brown, Les 24n.4
 bullshit job(s) 8, 16
 bureaucracy(-ies) 172, 177–80
 Burwell v. Hobby Lobby Stores, Inc. 189–90
 Buss, Sarah 202–3
 Busy Trolley 94–6
 Busy Trolley Three 95–6
 Busy Trolley Two 95
 Bykvist, Krister 95n.27
- Caney, S. 138n.10
 capitalism 4, 18–19, 35n.17, 56
 Carastathis, Anna 35n.17
 Cartwright, Nancy 168–9
 Cavanagh, M. 139
 Celentano, Denise 16, 16n.7, 19n.9
 Celikates, Robin 24, 29, 35n.17, 40–3, 54–5, 57, 58n.29
 Chamberlain, James 17–18
 Chan, J. 126, 138n.10
 Christiano, Thomas 199, 201–2, 222–3
 Citizens to Protect Overton Park 173
 Civil Rights Movement (CRM) 39, 44
 Clean Air Act 188–9
 Coase, Ronald 170
 Cohen, G.A. 253, 254n.15
 Cohen, Joshua 102n.2
 Collins, Patricia Hill 28n.9, 43, 47
 Collins, Stephanie 58n.29
 Combahee River Collective 44–5
 community 4–5, 8–9, 41, 51, 129–30, 232–3
 political 102–3, 103n.4, 104, 113
 compensation 82, 92–3
 consciousness 24–5, 24n.4, 28n.9, 42–3, 43n.20, 48, 198
 critical 42–3
 oppositional 46–9, 54–5
 raising 24–5, 24n.4, 43–4, 43n.20, 44, 44n.21, 48n.24, 49–54, 56–8
 consequentialist(s) 68n.2, 90, 125–6
 constitution 105n.8, 163–4, 180–4, 202–3, 208–9, 235n.6, 260–1
 contractualist 155
 Cordelli, Chiara 217n.31, 223n.39
 Couto, A. 67
 CR *see* consciousness raising
 Crary, Alice 43–4, 54, 58n.29
 critical theory 24, 24n.3, 29–31, 38–42
 Crow, Barbara A. 43–4
 Cruft, Rowan 263n.17
 Cruz, Joseph 25n.5
 cultural technē 23, 26–9, 35–6, 57–8
- Daniels, Norman 178–9
 Darwall, Stephen 106, 106n.10, 203, 242n.12
 Davis, N.A. 67, 78n.14, 83
 Dawson, Michael C. 35n.17
 Deaton, Angus 168–9
 Debreu, Gérard 169–70
 Defensive Shark 82
 Defensive Trolley 85, 88–90, 94–5
 democracy, deliberative 154
 democratization of learning 6
 deontic 228–9, 228n.3
 deontological 90, 151
 deontology 125–6
 distribution branch 177–9
 Dotson, Kristie 50–1
 DuBois, W.E.B. 47
 Duflo, Esther 168
 Dworkin, Ronald 34–5, 105n.7, 141, 157, 208n.18
- Eberle, C. 154n.19
 egalitarian 163–4, 176–8, 194–5, 197, 199, 205n.14, 221n.37, 223, 253
 elasticity 161–3, 178–9, 187–8
 Ellison, Ralph 47
 Enoch, David 110, 113n.29, 153n.17, 185
 epistemology 11, 24–5, 25n.5, 28–30, 43–4
 Equal Canoe 74–7
 equilibrium model 171, 184–5

- Estlund, David 58n.29, 100, 100n.1,
101–4, 106, 108, 108n.14, 213n.27,
257n.16, 263n.17
- Etchemendy, Max 223n.39
- ethics 3, 111–12, 129–30, 232–3
- Etninson, A. 120n.45
- Ewick, Patricia 43
- externalist 23n.1
- Fabre, C. 66–7, 77–8, 78n.14
- fairness 16, 128, 132, 132n.7, 133–4,
139–40, 155, 157
justice as 227–9, 228n.2, 229–36, 249,
251–8
- Fanon, Frantz 47
- Fard, Sahar Heydari 58n.29
- Feinberg, Joel 234n.5
- Fellow Passengers on a Train 239
- Ferguson, Ann 58n.29
- Finlayson, Lorna 24n.2, 58n.29
- Firth, J.M. 70
- Fishkin, James S. 186
- FitzPatrick, W. 94n.25
- Foot, P. 69
- Forst, Rainer 58n.29
- Foucault, Michel 9–10, 26, 46
- foundationalist 53–4
- Frankfurt School Critical Theory 24n.3,
38, 40–1, 57
- Frankfurt, Harry 107n.12
- Franklin, Benjamin 19n.8
- Fraser, Nancy 37–8, 50–1
- Frayne, David 14–15, 19n.9
- Friedman, Alexander 94n.25
- Friedman, Milton 177–8, 187
- Frye, Marilyn 43–4
- Galston, W. 154n.20
- Gardner, John 121n.46
- Garston, Bryan 223n.39
- Gaus, Gerald 114n.31, 153, 165n.2,
180n.9, 185, 185n.11
- Geuss, Raymond 38
- G. I. Bill 6–7
- Giddens, A. 27n.8
- Gifford, Daniel J. 175n.7
- Gilbert, Margaret 196–7
- Glendon, M. A. 120n.45
- Goldin, C. 13
- Gomberg, Paul 18
- Goodin, Robert E. 181, 219–20
- Gooding-Williams, Robert 58n.29
- Gosepath, Stefan 58n.29
- Graeber, David 8–9
- Hackett, Elizabeth 58n.29
- Hacking, Ian 9–10, 31
- Hampton, Jr., J. W. 188
- Hayek, F. A. 179–80
- healthcare 161–3, 174–5, 178–9, 183,
187–90
- Hegel 3–4
- Hobbes, Thomas 196–7
- Hodgson, L.-P. 103n.3
- Honneth, Axel 3–4, 40–1
- Hurka, T. 71–2, 77–8, 83
- idealization 163–4, 184–6
- ideology 23–32, 34n.15, 36, 38–43,
48–51, 54–8
- individualistic 231–2
- individuality 5
- Industrial Revolution 1–2, 4–5, 18–19
- Intemann, Kristen 28–9
- interpellation 25–6
- Iuliano, Jason 189n.14
- Jaeggi, Rahel 40–1, 58n.29
- James, Daniel 58n.29
- judgement 102n.2, 113n.29, 128–9,
138–40, 147–8
burdens of 109–10
liberal 144n.14
negative 147–50
non-perfectionist 138–9
perfectionist 138–40, 144n.14
privileged 110n.19
- judgment 30–2, 111, 113n.29, 169–70,
173, 188–9, 195, 201–3, 218
burdens of 164–5, 167–70, 235–6

- judgment (*cont.*)
 moral 35n.16
 normative 35n.16
 theoretical 30–1
 Jugov, Tamara 58n.29
 Julius, A. J. 221n.37
 justification 25n.5, 30–1, 40, 53, 83,
 101–2, 104–5, 113n.30, 116–17,
 141n.13, 144–5, 150, 198, 202–3,
 219–20, 241–3, 245–50, 262–3
 “civic republican” 175
 consequentialist 90
 empirical 53–4
 for harming 84–94
 lesser-evil 82–5, 88–90, 89n.21,
 90–5
 liability based 83
 moral 235–6, 241–2, 262–3
 objective 85, 90
 political 100–1, 102n.2, 105–6,
 109n.17
 problem 39
 public 120–1, 124–5, 144, 152–8
 Jütten, Timo 58n.29
 Jönch-Clausen, Karin 167n.4

 Kagan, S. 82–3
 Kant 103n.3, 209n.20, 212n.26
 Kantian(s) 102–5, 105n.6, 110n.19,
 118–19, 204n.11, 209n.20, 212n.26,
 213–14
 Kappel, Klemens 167n.4
 Katz, Lawrence F. 13
 Katzenstein, Emily A. 35n.17
 Keynesian 178
 Khader, Serene J. 28n.9, 34, 43, 51,
 58n.29
 Kolodny, Niko 200n.6, 206, 207n.16,
 211n.23, 212–13, 213n.27, 220n.34
 Koslow, Alison 58n.29
 Kramer, M. 129–30, 144n.14, 150–2,
 152n.16
 Kremer, Michael 168
 Krugman, Paul 186–7
 Kuhn, Thomas 54, 169

 Kukla, Quill 58n.29
 Kymlicka, Will 34–5

 Laden, Anthony Simon 228n.3
 Landemore, Hélène 223n.39
 Landis, James 175
 Larmore, Charles 105n.6, 124n.1, 154, 165
 Laurence, Ben 223n.39
 Lazar, Seth 67, 77–8, 83, 90–1, 91n.23,
 92–3
 legitimacy 55, 105n.8, 112n.26, 235n.6
 Democratic 260–1
 moral 151
 political 235n.6, 260–1
 Leland, R. 106n.9
 Lepold, Kristina 58n.29
 lesser-evil 82–5, 88–90, 89n.21, 92–6
 Lethal Violinist 88–9
 LGBTQ(+) 11, 52–3
 Liao, S.M. 120n.45
 liberalism 180n.10, 208–9
 classical 179–80, 180n.10
 comprehensive 101–4, 103n.3,
 106–9, 114, 117–19
 justificatory 180n.10
 neutralist 147
 perfectionist 110
 political 100–6, 105n.6, 108–10,
 115–17, 116n.37, 119n.42, 120–1,
 138n.10, 155, 157, 235–6
 public reason 154
 libertarianism 179–80, 210
 Lindemann Nelson, J. 73n.10
 List, Christian 196–7
 Lister, Andrew 165n.2
 Locke, John 210–11, 212n.26
 Longino, Helen 54n.28
 Lott, M. 154n.20
 Luddites 12
 Lugones, Maria 47
 Lunch Counter 259–61
 Lunch Queue Rule 260

 Mackinnon, Catharine 43–4
 Manne, Kate 51

- Mansbridge, Jane 46–8, 50–2
Maritain, Jacques 120, 120n.45
Marmour, Andrei 58n.29
Marx 3–4
McGeer, Victoria 27
McIntosh, D. 5
McMahan, Jeff 82–3, 90, 128n.4,
157n.22
McWeeny, Jennifer 43–4
Meazell, Emily Hammond 173–4
Medicare-for-all 183
metaphysics, moral 30
Meteor 90, 92–3
MeToo movement 45, 51
Metz, T. 126
Mill, J.S. 11–12, 117n.38, 118–19
Miller, David 131n.6, 157n.22
Mills, Charles 31n.11, 51
Moi, Toril 31n.11
Moody-Adams, Michele M. 53–4
morality 26, 52, 89n.22, 116, 209
natural 210–11
political 141, 141n.13, 155
sexual 138–40
Morris, Aldon 47–8, 50–2
Mother 71–3
Motor Vehicle Manufacturers
Association v. State Farm Mutual
Insurance Co. 173
Müller, Mirjam 58n.29
Muñoz, Cristian Pérez 223n.39
mutuality 233

Nagel, Thomas 217n.30
Neo-Keynesians 170
Neo-Nazis 52–3
Neufeld, Blain 228n.3, 263n.17
neutralist(s) 127–8, 140–1, 141n.13,
142, 147–50, 155, 157
neutrality 127, 132
liberal 132, 140–1, 144,
152n.16, 157
neutrality-perfectionism debate 127
New Deal 175, 189n.14
Nietzschean 135, 138

Nili, Shmulik 223n.39
nonnormative 52
normative 1–2, 10, 12, 15–16, 23–4,
29–31, 34n.15, 35–6, 52–8, 110–11,
155, 167, 202–3
autonomy 202–19, 222, 230–1,
234–7, 239, 241–2, 246, 248–9
challenge 24, 29
political epistemology 24–5
social theory 32–3
Nozick, Robert 80–1, 131–2, 179–80,
210, 210n.21, 212n.24, 220n.34
Nozickian:
minimal state 191
night-watchmen state 179–80
Nussbaum, Martha 37, 103n.3, 105n.6,
108–9, 112n.27, 115n.34, 117–18,
118n.39, 119–20, 120n.44, 154n.18

Oberman, K. 70
objectivity 33, 54
ontology 110n.20
social 1, 10
Ostrom, Elinor 32n.14
Owens, David 248–9

Pallikkathayil, J. 110n.19
Paradise Island 229–30, 256
Parental Shark 71
Pareto 169–70
paternalism 40–1, 132n.7, 147–50
Patten, A. 124n.1
Patterson, O. 11
perfection 135–7, 227–8
permissibility 69, 82–5, 90, 94–5, 124–5,
141n.13, 143, 210n.22, 216–17,
245–6
Pettit, Philip 196–7, 201nn.7–8,
204n.11, 205, 207n.17,
209nn.19–20, 212n.25
Philistinism 5
Plato 194–5
Platonic Aristotelian 194–5
pluralism 164–5
reasonable 144, 152, 155, 164–5, 170

- Pogge, T. 82–3, 103n.3
 Pohlhaus, Jr., Gaile 48, 58n.29
 Pollock, John 25n.5
preferential saving 68
 proportionality 77, 81

 quidnunc 144–5, 150–2
 Quiet Car Rule 246–7, 250, 259–60
 Quong, Jonathan 58n.29, 66–7, 70,
 70n.6, 71–2, 77–8, 81, 83–5, 84n.17,
 85–6, 86n.18, 87–9, 87n.20, 89–90,
 89n.21, 102–3, 105n.6, 113n.29,
 116n.37, 124n.1, 125n.2, 127, 133,
 141, 141n.12, 141n.13, 144,
 144n.14, 145–50, 152n.16, 156–7,
 157n.22, 165–7, 180, 263n.17

 Radzik, Linda 242n.12
 Railton, Peter 52
 rationality 185, 231–2
 Rawls, J. 10–11, 11n.6, 16, 34n.15,
 100–1, 104, 109–10, 118n.39,
 124n.1, 127, 135–9, 136n.9, 139–41,
 141n.13, 154n.20, 155–7, 161–72,
 176–80, 189n.13, 227–8, 227n.1,
 228–9, 228n.2, 229–33, 233n.4,
 234–6, 235n.6, 247, 249, 251–6,
 258, 262–3
 Raz, Joseph 17, 101, 110, 111n.21,
 113n.29, 127n.3, 145–6, 154n.20,
 205, 207n.17, 208, 209nn.19–20
 reasonableness 155–7, 156n.21, 157
 reciprocity 233, 233n.4, 235–6, 235n.6,
 247, 260
 Rector, Tully 58n.29
 Redstockings 43–4
 Reith, John 6
 republican 195–6, 204n.10, 213–14
 Rescue Sharks 68–9, 75
 responsibility 1–2, 6–7, 72–3,
 119–20, 234
 civic 6–7
 moral 81
 personal 134
 -sensitive 134

 Restrepo, P. 13
 Richards, J. 139
 Richardson, Henry 175, 189n.13
 Ripstein, Arthur 204n.12, 205n.14,
 207n.17, 209, 210–11, 212n.26
 Robeyns, Ingrid 228–9
 Rodin, D. 90
 Rose, Julie 223n.39
 Rosen, Gideon 58n.29
 Rössler, Beatte 16–17
 Rousseau, Jean-Jacques 195, 200,
 210–11, 222
 Rowena and Colin public reason
 example 165–9, 171, 182–5,
 189–90

 Sage, Daniel 17–18
 Salanie, Bernard 177–8
 Sandel, M. 138n.10
 Sarachild, Kathie 24n.4
 Saunders-Hastings, Emma 223n.39
 Scanlon, T.M. 34n.15
 Scheffler, S. 71–2, 77–8, 82–3
 Scheman, Naomi 47, 54, 58n.29
 Schouten, G. 124n.1
 Schwab, K. 1n.1
 Schwenkler, John 263n.17
 Scott, James C. 43
 Seidenfeld, Mark 175
 Self-Help Sharks 67–8, 74–6, 82
 Sen, Amartya 255–8
 Seo, Sarah A. 35n.16
 “separateness of persons”
 objection 231–2
 Sepinwall, Amy 263n.17
 Sewell, Jr., William 27n.8, 35n.17
 Shelby, Tommie 39n.19, 58n.29
 Sher, G. 126, 142
 Shiffrin, Seana Valentine 204n.12
 shit job(s) 8–9
 Siegel, Jonathan 189n.12
 Silbey, Susan S. 43
 Simmons, A. John 256–7
 Sinclair, T. 110n.19
 Smith, Adam 37

- Sobel, David 58n.29
 Socrates 194
 Solnit, Rebecca 36
 special others 66–7, 71–2, 72n.8, 73,
 77–8, 81, 83, 84n.17, 90, 96–7
 Speedy Shark 76–7
 Spiekermann, Kai 181
 stabilization branch 177–8
 Stahl, Titus 40–1, 58n.29
 Stemplowska, Zofia 112n.26, 121n.46,
 227n.1
 Stilz, Anna 198, 210–11, 223n.39
 Stoppenbrink, Katya 58n.29
 Stout, Jeffrey 27n.8, 58n.29
 Strangers on a Train 236
 Swift, Adam 112n.26, 227n.1
- Tadros, Victor 74–5, 94n.25, 94n.26,
 113n.29
 Táíwò, Olúfémí 263n.17
 Taurek, J. 69n.5
 Taylor, Keenanga-Yamahatta 44n.22
 Terlazzo, Rosa 97n.28
 Thompson, E.P. 5, 18–19
 Thompson, Marshall 263n.17
 Thomson, Judith Jarvis 69, 84, 94n.25
 Toole, Briana 43–4
 transfer branch 177–80
 transitivity 93
 Tremain, Shelley 37–8
 Tuana, Nancy 27–8
 Tuck, Richard 201
 Tuomala, Matti 177–8
 Two Meteors 91–3
- Unequal Canoe 75–7
 utilitarianism 230–2, 233n.4, 235–6, 252
- Valentini, Laura 227n.1
 Vallier, Kevin 165n.2
 van der Vossen, Bas 263n.17
 Vermeule, Adrian 176, 186
 Viehoff, Daniel 201, 211n.23,
 213n.27
 Violinist 84–90
 von Redecker, Eva 58n.29
- Wald, Patricia 173
 Waldron, Jeremy 110n.18, 116n.36
 Walen, A. 94n.25
 Walker, Margaret Urban 31–2, 54
 Wall, Steven 109n.16, 116n.35, 127–9,
 130n.5, 131–3, 141n.12, 155,
 157n.22
 Wallace, R. 219n.33
 Walton, G. M. 49n.27
 Wasserman, D. 94n.25
 Weber, Max 18–19, 19n.8
 Weinberg, Rivka 72–3, 73n.9
 Wendt, F. 153n.17
 Whitman v. American Trucking
 Associations, Inc. 188–9
 Whittington, Keith 189n.14
 Williams, Andrew 254n.15
 Women's Liberation Movement 44
 Wood, Allen 108–9
 Wright, Eric Olin 53
- Yablo, Stephen 58n.29
 Young, Iris 31, 46
- Zago, R. 12–14, 19n.9
 Zawidzki, Tadeusz 27
 Zheng, Robin 58n.29
 Zwolinski, Matt 179–80