UNIVERSITY OF CALIFORNIA, SAN DIEGO

Studies in Ideal and Non-Ideal Theory

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

in

Philosophy

by

Amy Elizabeth Berg

Committee in charge:

Professor Richard Arneson, Co-Chair
Professor David Brink, Co-Chair
Professor Saba Bazargan
Professor Gerry Mackie
Professor Donald Rutherford

2015

The Dissertation of Amy Elizabeth Berg is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Co-Chair

University of California, San Diego

2015

DEDICATION

To my parents, who taught me to read; to Debi Day, who taught me to love reading; to Alex Rajczi, who taught me to love reading philosophy.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I am enormously grateful to my co-chairs, Richard Arneson and David Brink, without whose exhaustive feedback and patient support this dissertation would never have become a reality. I am also thankful to Saba Bazargan, whose independent study on this topic got me started on thinking about many of the ideas in this dissertation. I have benefited from conversations with many of the other philosophy and political science faculty at UCSD, including (but not limited to!) Gerry Mackie, Kerry McKenzie, Dana Nelkin, Sam Rickless, Don Rutherford, Clinton Tolley, Eric Watkins, and David Wiens.

I have been lucky enough to talk over my ideas with some incredible philosophers, both at UCSD and in the wider philosophical community. Craig Agule probably knows some of the arguments of this dissertation better than I do. I also owe a lot to (among many others) Brian Berkey, Jeffrey Brand, Colin Chamberlain, Kathleen Connelly, Cory Davia, Gil Hersch, Tim Jankowiak, Kathryn Joyce, Alex Marcellesi, Noel Martin, JP Messina, Per Milam, Charlotte Newey, Theron Pummer, Erick Ramirez, Ben Sheredos, Mike Tiboris, Brian Tracz, Julie Walsh, Danny Weltman, and Alex Worsnip.

Over the course of graduate school, I received two fellowships that allowed me to concentrate on writing my dissertation: the US Department of Education's Jacob K. Javits Fellowship and the American Association of University Women's American Fellowship. I am grateful to these organizations for their generosity.

Finally, I am thankful to my family, who are a constant source of support and jokes about vegetarians. And to Matt Braich, who has made everything about writing a dissertation so much better.

Chapter 1, in part, is currently being prepared for submission for publication (as “Incomplete Ideal Theory”). The dissertation author was the primary investigator and author of this paper.

Chapter 2, in part, is currently being prepared for submission for publication (as “Ideal Theory and ‘Ought Implies Can’”). The dissertation author was the primary investigator and author of this paper.

VITA

2008	Bachelor of Arts, Claremont McKenna College
2013	Master of Arts, University of California, San Diego
2015	Doctor of Philosophy, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Studies in Ideal and Non-Ideal Theory

by

Amy Elizabeth Berg

Doctor of Philosophy in Philosophy

University of California, San Diego, 2015

Professor Richard Arneson, Co-Chair
Professor David Brink, Co-Chair

The distinction between ideal and non-ideal theory is a tool that can help us understand how to make moral and political progress. Ideal theory provides a goal for us to reach, and non-ideal theory tells us what to do in our current, non-ideal state. Throughout my dissertation, I argue that we need both of these kinds of theory in order to make progress. I also argue that we need to apply these tools to particular

problems in order to get a better understanding of the theoretical questions at stake. To that end, I investigate three particular problems. Chapter One is devoted to showing that we need ideal theory to make sustained societal progress over time. But because we are unable to agree on a complete ideal, we should work together to create incomplete ideal theory, which can then guide our progress. Chapter Two shows how we can use the ideal/non-ideal distinction to resolve two longstanding tensions in moral and political philosophy. We disagree about how much our moral theories should yield to our flaws, and we also disagree about how to interpret the voluntarist constraint: what it means for "ought" to imply "can." I show that we need ideal theory of morality, which uses a thinner version of the voluntarist constraint and does not yield to our flaws, to provide an ultimate standard. But we also need non-ideal theory, which uses a thicker version, to guide our actions. Chapter Three tackles beneficence. Does our duty to the very poor increase when others inevitably fail to comply with that duty? It may be that we only have to do our fair share—that even in the non-ideal world, we only have to do what we would have had to do in the ideal. I show that this view is plagued by counterexamples. Many consequentialists hold the alternative view, that we must pick up others' slack, but their interpretation of this view is extremely demanding. I argue that we should look to an alternative moral theory. Two versions of deontology, intuitionism and Kantianism, require us to do more when others are doing less without also making extreme demands.

Introduction

In *A Theory of Justice,* Rawls mentions, almost in an aside, that his two principles are an "ideal theory" of justice: something that we should try, but may be unable, to achieve (215). "Non-ideal theory," on the other hand, tells us what to do in the world that we live in right now. Rawls issues a kind of challenge: his ideal theory won't tell us everything about what to do in the real world, so we must develop a non-ideal theory as well, a task he doesn't take on here. Although Rawls himself makes repeated references to this distinction between ideal and non-ideal theory in his later work, it was all but ignored in favor of other parts of his work.

That changed about ten or fifteen years ago. Suddenly political philosophers were animated by what turns out to be a deep and important question: what should the methodology of political philosophy be? One possibility is the one Rawls presents: we should start by trying to identify an ideal state and then work our way there. We might think of it as the standard approach in political philosophy historically, going all the way back to Plato's *Republic*. But a challenge has arisen from philosophers who are proponents of non-ideal theory to the exclusion of ideal theory. Anderson, Sen, Mills, and others argue that the emphasis on starting from an ideal has fundamentally distorted political philosophy. If the ideal is distant, perhaps impossible, surely it's a waste of time to start with the ideal. Perhaps it even sets us back, if the search for the ideal distorts our vision of what is possible or desirable for us today. It would surely be better, these non-ideal theorists argue, to start with more facts. Start with what we know about our non-ideal world: our actual political and economic systems,

individuals as they actually are, social ills we know we want to cure. Rather than doing pie-in-the-sky philosophizing, we should figure out how we can cure those social ills, and then move forward from there.

And as this debate between ideal theorists and their critics has blossomed, the ideal/non-ideal theory literature has spread into other areas as well. Political philosophers are hard-pressed to agree on what ideal (or non-ideal) theory even is, so a literature has flourished on the definitions of the terms (see, for example, Valentini, "Conceptual Map"). Many questions of what ideal theory should do turn on what we mean when we say that something is "feasible," so some philosophers working on ideal and non-ideal theory have turned their attention to issues of feasibility (here see Gheaus; Gilabert and Lawford-Smith; Lawford-Smith; and Raikka, among others). And some philosophers contend that discussions of ideal and non-ideal theory can't be divorced from the contexts in which we use the terms, so we now have this distinction as it's applied to integration (Anderson) or to global justice (Ypi).

0.1 Theoretical commitments

Into this complicated, messy, and fascinating landscape comes my dissertation. Here are a few of the theoretical commitments I have and will attempt, to varying degrees, to defend: First, I believe that ideal theory is necessary for making certain kinds of progress, although specifying how it is necessary, and what kinds of progress, is a challenge I will take up in Chapter 1. Ideal theory can, when done correctly, serve as a map that can help us chart our progress. Perhaps even if we never reach the ideal, knowing what the ideal is and why it's ideal can help us to figure out what

improvements to make next. Perhaps ideals aren't necessary in all cases; perhaps they aren't even helpful in all cases. But I think there are good reasons to reject the blanket assertions of some non-ideal theorists that ideal theory is always unnecessary or unhelpful.

Second, I believe that the nature of ideal theory is context-dependent. Ideal theory within a particular branch of moral philosophy may be different—be derived differently, have different content, be applied differently—than ideal theory somewhere else in moral philosophy, or in political philosophy. In the same way, normative ideals will function differently in different cases, depending on what work we need them to do. The ideal society provides very limited action guidance for us right now. It presents an ideal, to which we add facts about our actual world to get non-ideal theory, but it might be disastrous for us to implement Rawls's two principles right now. We should think of this ideal as providing an ambitious but long-term target. By contrast, other ideals hit much closer to home. Virtues are a kind of ideal, maybe even an impossible one, but whatever the virtues are, they should guide the actions we take right now.

This means that there may be only a limited sense in which there is such a thing as a single correct conception of ideal theory. One problem with the literature on ideal and non-ideal theory thus far is that it has tended to treat the two halves of this distinction as two monoliths. If "ideal theory" in the case of justice entails full compliance, then "ideal theory" in general must entail full compliance. This has probably arisen for reasons that are at least partly genealogical. Because the ideal/non-

ideal distinction arose within the context of Rawlsian political philosophy, ideal theory has tended to mean theory about the ideally just society. When philosophers consider ideal theory in the abstract, they have sometimes ported over assumptions about this kind of ideal theory into ideal theory more generally. This has made it difficult to consider ideal theory as an abstract entity, because the features of ideal theory of justice may not be part of ideal theory about morality. This has led to confusion about what ideal theory actually is. While I think the theoretical debates are important, I think it might be time to try a different tack. It would serve us well to think about ideal theory in relation to specific problems in moral and political philosophy, as I do in Chapters Two and Three. Using ideal theory here can help clarify some of the issues in these persistent and ongoing debates on topics such as the demandingness of morality (Chapter Two) and our duties to the very poor (Chapter Three). But it may also help to clarify the theoretical landscape. Once we see what different kinds of ideal theory can do, we may be in a better position to figure out what, exactly, it is. Doing the theory before figuring out its applications hasn't seemed to work so far: we should start applying it more, and then we can see the productive interplay between theory and application.

Third, in spite of this, I believe that there is such a thing as an ideal/non-ideal distinction. There are at least some unifying features of normative ideals, and it can be useful to consider how various problems have ideal and non-ideal components. These ideal and non-ideal theories may not share all of the same features, and they may be put to different purposes. But it will help us to consider what they share. For that

reason, I will use a relatively ecumenical definition of "ideal theory" throughout this dissertation. I discuss that definition below.

And finally, in concert with what I've said above, I believe that ideal theory is just as relevant and important in moral philosophy as in political. As I have said, these problems have arisen in political philosophy for genealogical reasons. But moral philosophers must also consider methodological questions about how moral philosophy ought to be done. Should we do moral theory for angels and then try to apply it to humans, or should we start with human flaws and then try to build moral theory around them? There is no reason to confine ideal theory, either in definition or in application, to the search for the ideally just society; there is no reason to think of non-ideal theory as specifically concerned with alleviating societal problems. A large portion of my dissertation is devoted to ideal theory within moral philosophy, both as a way to illuminate long-standing problems in moral philosophy and as a way to rethink the ideal/non-ideal distinction generally. This will improve our understanding of moral philosophy, but it will also build links to political philosophy. For example, I argue in Chapter Two that the distinction between normative, realistic, and moderate moral theories is mirrored by a distinction between normative, realistic, and moderate political theories. The result I come up with in the moral case thus has ramifications for the political case.

0.2 The ideal/non-ideal distinction

I've said some things about why I think ideal theory can be valuable and about some of the commitments I defend throughout this dissertation. It's now time for me

to say something about what the ideal/non-ideal distinction is. There are a number of different versions of this distinction that have been advanced; Laura Valentini lays out many of them ("Conceptual Map"). Ideal theory might be theory that assumes full compliance, while non-ideal theory assumes that people will only partially comply (Valentini, "Conceptual Map" 655-56). Of course, then we need to figure out what we mean by full compliance (Eyal). Or ideal theory might be theory that is (relatively) fact-free, while non-ideal theory takes more facts into consideration (Valentini, "Conceptual Map" 656-60).

As I suggested above, I think relatively narrow definitions such as these can sometimes cloud the debate. If we expect all ideal theory to involve members of a society who comply fully with the principles of justice, then that leaves out ideal theories where full compliance is not a central feature. For example, I say nothing in Chapter Two that particularly has to do with full compliance. There, I distinguish ideal from non-ideal moral theories by which version of possibility they use in determining the obligations we have. This says nothing about which obligations we actually comply with. Since full compliance has generally had to do with what we can expect other members of our society to do, it seems more relevant to ideal theory of justice than to this project within ideal theory of morality. On the other hand, Chapter Three is concerned specifically with how our obligations of beneficence are affected by whether others are fully or partially complying with their duties.

Similarly, Chapter Two bears some relationship to the fact-free/fact-using conception of the distinction. In order for non-ideal moral theory to give us action

guidance, it must take facts about our psychological and motivational weaknesses into account. A theory that takes more facts into account can give us more specific and usable action guidance. But Chapters One and Three do not rely on this distinction. In Chapter Three, the distinction between ideal and non-ideal beneficence is a distinction in how many people comply with the duty of beneficence, not in what we know about (for example) people's needs, reasons for giving or not giving, or the causes of poverty.

Thus, the general definition I will use is this: *Ideal theory tells us about the best version of something. Non-ideal theory tells us what to do when we aren't in the ideal (whether because we won't or because we can't).* The ideal could be anything—the fully just society, ideal morality, perfect virtue. It doesn't have to be the best *possible* version of something—depending on what the subject matter is, the ideal might be impossible (say, in the case of perfect virtue). The ideal doesn't have to be fleshed out. Ideal theory may simply present an ideal without giving reasons behind it, although better ideal theories usually will explain why the features of the ideal are ideal. One way to head off Mills's critique that ideal theory is too focused on what upper-class white men think is ideal is to consider what the reasons are for holding those ideals, to see if they are truly biased. A good ideal theory will be open to this kind of interrogation because it will make plain why the elements of the ideal are ideal.

There is a distinction within non-ideal theory that will frequently be relevant, although it is not essential to the definition of non-ideal theory. This is the distinction

between *transitional* and *non-transitional* non-ideal theory. Some non-ideal theory takes as its concern how we can get from our current situation to the ideal. Where Rawlsian ideal theory is concerned, for example, we might ask how we can get from a society not governed by the two principles, where some people fail to comply with the principles, and where background conditions are unfavorable, to a situation in which everyone can and does comply with the two principles. This is a question of transition, and transitional non-ideal theory can guide us here. On the other hand, we need non-transitional non-ideal theory as well. Even in cases where punishing or aiding people gets us no closer to the ideal (the criminal whose lengthy trial inspires copycats), we still might think appropriate to do so. There are some cases in which non-transitional non-ideal theory constrains our progress. In the case of ideal justice, we cannot institute the two principles by brutally repressing the human rights of everyone in the society. Non-transitional non-ideal theory requires us to have some respect for human rights (although how these two theories interact is an important separate question). As I say, this distinction is not part of the definition of non-ideal theory; we can do sometimes do transitional or non-transitional non-ideal theory on its own. The discussion of beneficence in Chapter Three is almost exclusively concerned with non-transitional non-ideal theory. But where one of the jobs of non-ideal theory is to direct our progress toward the ideal, this distinction will be important.

This definition of the ideal/non-ideal distinction gets at the same basic idea that Rawls's does. The two principles of justice are Rawls's ideal theory because implementing them will give us the fully just society. Non-ideal theory, while derived

from ideal theory, gives us different obligations. If the two principles of justice are out of our reach because our society is too poor, we will need to implement different principles of justice. So the focus in Rawls for ideal theory is laying out the best, and the focus for non-ideal theory is saying what to do when we can't get there.

But Rawls could have formulated his ideal theory in other ways. All we need for Rawls's ideal theory is full compliance with the principles of justice, but he could instead have assumed perfect virtue. In this ideal theory, individuals don't just comply with the principles of justice; they comply with every moral rule, are altruistic in the extreme, and so on. This would be a different kind of ideal theory, but it too would be compatible with the definition I offer here.

This definition is also compatible with ideal and non-ideal theories that aren't explicitly Rawlsian or concerned with justice at all. In his discussion of consequentialist duties of beneficence, Murphy identifies the ideal of beneficence with full compliance—that is, everyone does her fair share to improve the position of the distant needy. He says nothing about the favorability of background conditions; indeed, since there are distant needy who require our aid, we might assume that background conditions are distinctly *un*favorable. But a society in which everyone does his fair share to help the distant needy is the best version in the relevant sense; a society in which some people fail to comply is non-ideal in the relevant sense. The ideal/non-ideal distinction I have articulated is compatible with this way of understanding ideal and non-ideal theory as well.

So this definition can cover many uses of the terms "ideal" and "non-ideal." It is sufficiently general to apply to different kinds of theorizing about justice as well as other kinds of ideal, while preserving what they all have in common—namely, a view of the ideal as the best version of something. For this reason I believe it's a useful way to understand ideal and non-ideal theory, even if specific theories differ in their conceptions of this broader concept. But I won't pretend that it captures every way people have made use of this distinction. For example, Colin Farrelly places ideal and non-ideal theory of justice on a continuum distinguished by their reliance on facts: more ideal theories abstract from the world as it is, while more non-ideal theories incorporate more facts (847). While I have no problem with the view of ideal theory as a continuum (see Chapter Two), Farrelly's view changes the focus of the distinction from normative (whether the theory describes the best version of something) to epistemic (how much we know about the world as it is). I think this focus on facts is a little bit of a red herring. It is true that often ideal theories will contain fewer facts, since they will describe what the ideal would look like, not the political and social conditions that exist today. They may not concern themselves with the ins and outs of particular human psychologies, out of a thought that concern with the facts about human weaknesses makes a theory less ideal (see Chapter Two here). But facts do not necessarily operate this neatly. Some non-ideal theory may be broad and abstract from particular facts. Some ideal theory may be quite detailed and may pay a lot of attention to facts about human nature. So while it's natural to see facts as going with non-ideal

theory, I don't think Farrelly's fact-based continuum is looking in the right place, nor does my ideal/non-ideal distinction capture it.

This may sound strange, given that Chapter Two displays this kind of fact-based continuum. When we move from ideal to non-ideal moral theory, we add in more facts about an individual—that person's psychological and motivational flaws. This allows non-ideal moral theory to provide the immediate action guidance that ideal moral theory cannot. But I think it would be a mistake to think that the relationship of ideal and non-ideal theory to facts is *definitional* of the theory laid out in Chapter Two. The ideal theory here still is the best version of our moral theory—what we ought to do if we have no psychological or motivational incapacities. The non-ideal theory still tells us what to do when we can't reach the ideal. Facts have a role to play, just not a fundamental one.

Another definition that is somewhat orthogonal to mine is Cohen's. What is thought of as Cohen's "ideal theory" consists of principles stripped of nearly all their content. Farrelly, for example, cites Cohen as being on one extreme of his ideal-to-non-ideal continuum) (847). Cohen makes a particular fact-principle argument for thinking that all of our principles bottom out in fact-free principles, but the details of this argument need not concern us here. What is important is that Cohen is often thought of as coming up with an extreme version of ideal theory (see Valentini, "Conceptual Map" 657). Cohen is supposed to have come up with ideal theory for the angels, theory that presumes perfection instead of Rawls's somewhat more modest claims about human nature.

But I think this misconstrues Cohen's project. Cohen argues that our most fundamental moral principles have *no* determinate content whatsoever and cannot guide actions for angels or for people. We must fill in these principles (for example, "absent other considerations, one should avoid causing pain") with whatever facts apply in the circumstances, whether we are talking about angels or about us (Cohen 245). Cohen's fact-free principles are ideal *only* in the sense that they do not tell us about particular circumstances, not in the sense of telling us what the best version of anything looks like. It seems to me that conflation between these two kinds of ideal has caused misinterpretation of Cohen's ideas in this area (and it doesn't help that he is an ideal theorist in other senses of the term as well). I think there are good reasons for keeping this conception of ideal theory distinction from ideal theory as theory about the best, and I think failure to keep it separate has caused confusion. When I talk about "ideal theory" here, I'm not talking about Cohen and the fact-principle argument.

I have laid out my version of the ideal/non-ideal distinction and discussed some differences between it and other views. For reasons I have given, I think my distinction better captures what we talk about when we talk about ideal and non-ideal theory than Cohen's or Farrelly's does. But that does not mean that the distinction I've laid out answers all questions about how we should think about ideal and non-ideal theory generally. I don't think we should think of the use of facts as constitutive of the distinction between ideal and non-ideal theory, but we might have legitimate disagreements about how many facts we need in order to know what the best is—

whether it's the best possible state of affairs or the best state of affairs within certain constraints. And there may be other fruitful ways of understanding the ideal/non-ideal distinction than those I have outlined in this dissertation. While a reasonably ecumenical account of the distinction can help us to see what all of these theories have in common, there are different ways to use ideal and non-ideal theory in moral and political philosophy. Thus there may be different fruitful ways to understand the distinction. I am only exploring a few here.

0.3 The plan of the work

I've given an overview of some of my guiding thoughts on ideal theory—what it is, what it isn't, why it's important. In what follows, I will outline what I plan to do in this dissertation.

Chapter One

In Chapter One, I engage most directly with the existing literature on ideal and non-ideal theory. As I said above, one of my theoretical commitments is that ideal theory is valuable, sometimes even necessary. In this chapter, I will explore why that is the case. I will do so by examining the Sen/Rawls debate on ideal theory. I begin by laying out Rawls's ideal theory. This comes largely out of *A Theory of Justice,* but I will also use Rawls's later work on the distinction, such as in *The Law of Peoples.* I attempt to come up with the most coherent and workable account of ideal and non-ideal theory and their various subparts.

I then turn my attention to Sen's criticisms of this and other ideal theory. Sen has three particular criticisms: that ideal theory is not sufficient to tell us everything

about what to do, that it is not even necessary, and that we cannot come up with a complete, or "totalist," ideal theory. The first criticism is easily dismissed; ideal theorists do not think that ideal theory alone can do the job, nor should they. The second criticism should give us more pause. I can agree with Sen that ideal theory is not necessary in all circumstances. This is one reason that, as I argued above, we should not see ideal theory as a monolith. But ideal theory is necessary in precisely the circumstances Rawls wants to use it for—that is, situations in which we want to make sustained societal progress over time.

But this brings us to the third criticism, that thanks to disagreement and incomplete information we may never be able to agree on an ideal. If it's necessary, but we can't do it, then what happens next? Taking a page out of Cass Sunstein's book, I argue for incomplete ideal theory. We should start by seeing what particular facts about the ideal we agree on. We can then build agreements around those facts without having to agree on the theory behind why those facts are ideal. As we go along, we revise our ideal and come to agreement on theoretical principles. Ideal theory of this kind can accept incompleteness while still guiding our progress.

Chapter Two

In Chapter Two, I extend ideal and non-ideal theory to moral theory. I begin with a well-known debate about what demands moral theory should make on us. Should it require us to do things that, while perhaps possible for more ideal versions of ourselves, are impossible for us given our flawed motivational and psychological structures? Or should it hew more closely to what we're really like? I argue that the

first position, as expressed by normative moral theories such as consequentialism, can be too demanding and psychologically unrealistic. But the other extreme, as found in Bernard Williams and Susan Wolf, is too lax and too complacent about what we're actually capable of. At the same time, moderate moral theories, such as Owen Flanagan's, fail to capture what's distinctive and attractive about the extremes.

There is a related problem. Philosophers have struggled to find the right interpretation of the voluntarist constraint, "ought implies can." In particular, is "can" expressing physical possibility, motivational possibility, or something else? I argue that all of these versions of possibility have their place. Just as physical possibility in general is not more "true" than psychological possibility, so there is no one true version of "ought implies can." Instead, we must develop ideal and non-ideal moral theories, in which the ideal version of a moral theory uses a very thin version of the voluntarist constraint and its non-ideal counterpart theories use thicker versions. This solves the problem of how much morality should yield to what we are like. The ideal version of a moral theory may be extremely unyielding, but we can recognize that its demands are not always possible for individuals. That is why we need non-ideal moral theories with thicker, non-ideal senses of "can." When the ideal's demands are out of reach, non-ideal moral theory can provide action guidance.

This is an argument within moral philosophy, but it has ramifications for political philosophy as well. I show how we can arrange political theories along a similar continuum according to the assumptions they make about human nature. This gives rise to another kind of ideal/non-ideal distinction within political theory besides

the standard Rawlsian one. Taken together with my argument in this chapter that some ideal theories exist on continuums with their non-ideal counterparts, the argument about political theory shows how considering applications of ideal and non-ideal theory can help us to make progress on the theoretical issues.

Chapter Three

In the third chapter, I consider whether we need non-ideal theory as well as ideal theory to govern our obligations of beneficence. This is a debate that has been of particular concern to consequentialists. Liam Murphy takes issue with Peter Singer's extremely demanding view of this duty. On Singer's view, we must increase our giving, perhaps substantially, in situations of partial compliance, when others are failing to do their part. If we do not, the very poor will suffer and die as a result of our noncompliance. Murphy argues that this is fundamentally unfair. Instead, all we owe is what we would owe in ideal situations of full compliance—we only have to do our fair share.

I weigh Murphy's reasons for holding this position and find them wanting. Murphy has trouble with cases of easy rescue: when you and another person are in a position to rescue two people, Murphy thinks you have no obligation to rescue the second person. So Murphy's position is implausible, but Singer's is extremely demanding. This inspires a search for an alternative account of beneficence. In Chapter Three, I show that deontology does a better job than these consequentialist theories do of telling us our duty of beneficence. I examine two deontological alternatives, Ross's intuitionism and Herman's Kantianism. Both of these theories

require us to pick up others' slack without saddling us with the extreme demands that Singer's view does. Deontology is thus an attractive moral theory to use in thinking about our duties of beneficence.

Conclusion

Finally, I return to the main themes I have outlined in this introduction—when ideal theory is necessary and helpful, what its relationship to non-ideal theory is, where there are connections between moral and political philosophy. I will bring out the commonalities that have developed among the three chapters of my dissertation, and I will consider where there are differences in the ideal and non-ideal theories of each. And finally, I will chart some future directions for research into ideal and non-ideal theory.

Chapter One

Rawls, Sen, and Incomplete Ideal Theory

In the middle of *A Theory of Justice,* Rawls takes a break from developing his theory of the just society to point out that the theory won't apply in all circumstances. While he's given us a theory for the ideal case, in which we're able to achieve justice and everyone complies with the requirements of justice, there are many non-ideal cases in which justice won't or can't obtain. Our current society is one of these, as are all other existing societies. Rawls calls his theory of justice "ideal theory," and he says that we need a companion "non-ideal theory," which will tell us what to do in non-ideal circumstances and how to reach the ideal.

Although Rawls briefly takes up the ideal/non-ideal distinction in his later work, particularly in *The Law of Peoples,* for the most part this was a relatively overlooked part of his theory. Recently, however, ideal and non-ideal theory have gotten more attention. Some of this attention has come from critics of the general project of doing ideal theory. While Rawls sees ideal and non-ideal theory as going hand-in-hand, these critics argue that ideal theory is useless (or worse), and we should focus instead on non-ideal theory. When our world is so clearly non-ideal, one version of this critique goes, what good does it do to think about the ideal? Amartya Sen has recently been at the forefront of this critique. If ideal theorists are to continue their project, they must find a way to answer or accommodate the critiques of Sen and others. This can be done, but it comes at a cost to ideal theory's claims to completeness.

In this chapter, that's what I aim to do. The chapter consists of three sections. In the first, I go through Rawls's discussions of ideal and non-ideal theory. I discuss the distinctions he makes between ideal and non-ideal theory and within each type of theory. I also discuss some puzzles for how to understand Rawls's ideal theory: the attainability of the ideal, the relationship between blameworthiness and non-ideal theory, and the relationship between ideal and non-ideal theory. The result is a reconstruction of perhaps the most significant ideal theory of justice.

This sets us up for the second section, in which I consider the critiques of ideal theory Sen makes in *The Idea of Justice.* Sen is right that ideal theory is not sufficient to tell us everything about non-ideal justice, but he is wrong to claim that ideal theory is unnecessary for guidance. We need ideal theory in order to make sustained societal progress over time; without it, we risk getting stuck at a dead end. Sen's most important and plausible critique of ideal theory is his claim that we will have to live with incomplete theories, due to biases, gaps in information, and disagreements.

The third section of this chapter is devoted to finding a way that ideal theorists can live with this incompleteness. I use Sunstein's discussion of incompletely theorized agreements, in which we come to agreement on particulars and then find overlap with each other in order to build consensus. This shows how to do ideal theory under non-ideal conditions of disagreement and incomplete information. We can come to incompletely theorized agreements about what the ideal would look like. This incomplete ideal will become more complete over time as we work to attain it. We must modify our expectations about ideal theory in order to arrive at an ideal that we

can use to guide our progress, even if it's incomplete. In the end, I attempt to take Rawls's most important insights about ideal theory and square these with the insights Sen has given us about the impossibility of complete ideal theory. There are significant areas of potential agreement between Rawls and Sen which I will make use of in figuring out what kind of ideal theory we can reasonably expect to give guidance about our actions.

1.1. Rawls

Although Rawls's distinction between ideal and non-ideal theory has spawned most of the current debate in the literature, untangling the distinction is not simple. I start here with a summary of what I believe Rawls intends his account of ideal and non-ideal theory to be.[1] I then look at the places Rawls develops that account—first in *A Theory of Justice* and later in *The Law of Peoples*—to create a more detailed picture of the account. The details change slightly over the course of Rawls's work, so the full picture of the account will only emerge when we consider how it evolves over time. After I present the best version of Rawls's account, I go into more detail about several puzzling features of the account as it appears in *A Theory of Justice*: whether all of ideal theory is attainable, the role of restrictions on liberty, how we should view the branch of non-ideal theory we're not blameworthy for, and what the relationship between ideal and non-ideal theory is. Finally, I flesh out the account by considering how ideal and non-ideal theory are portrayed in Rawls's later work, especially *The Law of Peoples.*

[1] Simmons takes on this task as well. Although our accounts are similar on many of the important points, he also raises different problems than I do, and we have some disagreements, which I highlight below.

Rawlsian ideal theory: a summary

Rawlsian ideal theory, as it emerges over time and throughout Rawls's works, is composed of a series of increasingly fine distinctions. First, we distinguish:

1) Ideal theory
2) Non-ideal theory

On the Rawlsian account, 1) ideal theory sets a standard or a goal for us to try to achieve—in the case of domestic justice, for example, ideal theory tells us the principles the just society is organized by, and we judge whether societies are just by comparing them with this ideal (Rawls, *Theory* 216).[2] In the case of international justice, ideal theory gives us the Law of Peoples, which tells us how societies must conduct themselves with respect to other societies in the world. In addition to telling us what the ideal is like, ideal theory also provides us with information that can be helpful in constructing its counterpart, non-ideal theory (Rawls, *Theory* 216).

2) Non-ideal theory describes what we must do in situations that aren't currently ideal (Rawls, *Theory* 215-6). It is itself divided into two parts:

2a) Unfavorable conditions

2b) Deliberate injustice

2a) is theory about unfavorable conditions are things that humans are not responsible for, can't be blamed for, or have no control over. This includes background conditions within a society, such as its level of natural resources, and events that aren't under anyone's control, such as natural disasters. 2b) is theory about anything that humans

[2] All references to *A Theory of Justice* are to the revised (1999) edition.

are responsible for, can be blamed for, or control, such as racism, sexism, or greed. The problems 2b) deals with include past injustices—for example, a society that has racial disparities due to past racism would count as an example of deliberate injustice, even if no one in this society holds racist attitudes anymore. But the exact distinction between 2a) and 2b) is fuzzy (more about this below).

Finally, in some places Rawls makes a third distinction, within 2a) (*Theory* 215). Rawls sometimes distinguishes:

2a1) Natural limitations and accidents of human life

2a2) Historical and social contingencies

We can see now that natural disasters and natural resources fit into 2a2), historical and social contingencies. These are features that make a society non-ideal but that don't have to do at all with human nature or action. In contrast, 2a1) is difficult to pin down. In *A Theory of Justice,* at least, Rawls seems to think that there are features of human nature that aren't deliberate injustices (that is, for which we're not blameworthy or responsible, and which we don't control) and that don't exist in the ideal. It's hard to say exactly what these are, and I discuss this below as well.

Finally, there's one last distinction within non-ideal theory, which Rawls hints at but which he doesn't explicitly make. This is the distinction between:

A) Transitional non-ideal theory

B) Non-transitional non-ideal theory

As Rawls lays it out in *A Theory of Justice,* non-ideal theory is essentially identified with A): he writes about non-ideal theory in terms of "meeting" and "removing"

injustice (*Theory* 216). This kind of non-ideal theory is directed at transitioning from our current non-ideal state to the ideal. But there's a second kind of non-ideal theory, B). B) is not directed at transitioning to the ideal; rather, it tells us about duties we have in non-ideal situations whether or not complying with those duties gets us closer to the ideal. Rawls's general conception of justice might be an example of B), and I take this question up below. To see the distinction between A) and B), consider our duties in the face of severe poverty. We might have A) transitional non-ideal duties to move toward an ideal in which severe poverty doesn't exist; these might include duties to alleviate the structural causes of poverty. But we might also have B) non-transitional non-ideal duties to relieve the suffering of the very poor, whether or not doing so gets us closer to the ideal. While Rawls emphasizes the role of transitional non-ideal theory (for simplicity's sake, "transitional theory") over non-transitional non-ideal theory ("non-ideal theory"), both are part of his non-ideal theory.[3]

[3] Simmons makes one final distinction, between different subject matters of ideal theory—basic structures, individuals, and nations (17). So his map of Rawls's ideal and non-ideal theory has nine parts: there is ideal theory, theory about deliberate noncompliance, and theory about unfortunate noncompliance (the last two correspond to deliberate injustice and unfavorable conditions) for each of the three subject matters. Unlike Simmons, I don't believe that this further distinction is fundamental to the structure of Rawls's ideal and non-ideal theory in the way that the previous distinctions are. Instead, once we have that structure down, we can move on to formulating ideal and non-ideal theory for specific subjects (not just domestic and international justice but other facets of moral and political philosophy as well).

Here is a map of these distinctions:

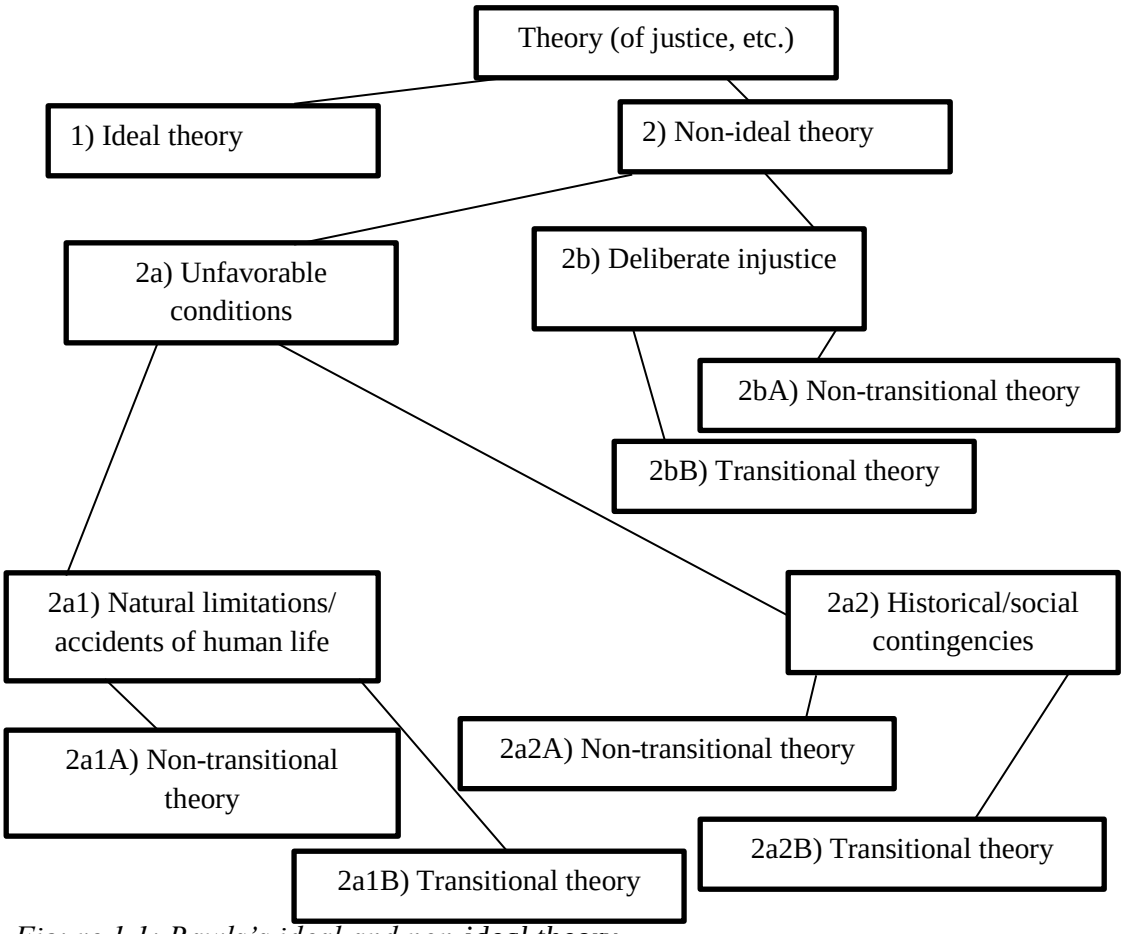

Figure 1.1: Rawls's ideal and non-ideal theory

Some questions about **A Theory of Justice**

This is the framework Rawls presents us with, but there are some puzzles that remain to be solved. It's not clear whether Rawls thinks this ideal theory can be achievable. Rawls initially defines non-ideal theory as a way to tell us when restrictions on liberty are appropriate, but it seems like there will be certain kinds of restrictions on liberty (such as on the liberty of children) even in the ideal world. And

looking at the general conception of justice can help us to understand the distinction between transitional and non-transitional non-ideal theory.

The attainability of ideal theory

One puzzle is whether Rawls believes that we can achieve the ideal. Sometimes it sounds like he does: he defines ideal theory as presenting "a conception of a just society that we are to achieve if we can" (Rawls, *Theory* 216). It's unlikely that Rawls would take the condition "if we can" to be unfulfillable; he must think there's some chance we can achieve the ideally just society. But consider what's required for that ideally just society. We must lack unfavorable conditions, including any natural disasters that would prevent us from instituting the two principles, and we must also have full compliance—that is, everyone in the society must fully comply with the principles of justice. It is this second condition in particular that seems unattainable. It seems easier to imagine a society with enough wealth to provide for everyone and protect against natural disasters than it does to imagine a society in which every person fully complies with the requirements of justice. Full compliance would require a radical revision in human nature. Does Rawls really think this is possible?

There's one suggestion that he does. Rawls writes that "Men's propensity to injustice is not a permanent aspect of community life; it is greater or less depending in large part on social institutions, and in particular on whether or not these are just or unjust. A well-ordered society tends to eliminate or at least to control men's inclinations to injustice…" (*Theory* 215). Rawls appears here to believe that our moral

psychology is plastic: if institutions are set up to be just and to require us to be just, we will become able to comply (more) fully with the requirements of justice. Thus, the full-compliance part of ideal theory is attainable.

Unless we have Rawls's rosy picture of human moral psychology, this may seem unrealistically optimistic. In response to this kind of problem, Andrew Mason has suggested breaking down ideal theory into different levels of analysis (265). On Mason's account, there are three levels: the first and most abstract elucidates the best *reasonable* principles for the ideal: these are principles that are not over-demanding, that people could live under. The second elucidates the best *feasible* principles: these are principles that would guarantee stability, perhaps in the face of historical constraints. The third elucidates how to balance that ideal against others: it balances justice against other things we care about. Mason charges Rawls with failing to distinguish between these levels of analysis (268). Given this understanding of ideal theory, the two halves of Rawls's ideal justice might occupy different levels of analysis. Full compliance occupies the first level, because it's reasonable to demand of people that they comply with justice, but it's not feasible to expect that they will actually do it. On the other hand, the removal of unfavorable conditions occupies the second level; it is reasonable, and many parts of it seem to be potentially feasible as well. If we don't accept Rawls's view of moral psychology, then Mason's levels of analysis can help to explain the mistake Rawls has made. Unfortunately, they also suggest that the two parts of Rawls's ideal theory are not equally attainable. On a revised theory, then, Rawlsian ideal theory still has two parts, but they're treated

somewhat differently. The favorable-conditions half is attainable; the full-compliance half is not. Yet full compliance can still guide our actions. We can structure our institutions to promote compliance with justice, and we can hold people responsible when they fail to live up to the demands of justice. Full compliance can thus guide our actions even if we never achieve it.

Restrictions on liberty

In *A Theory of Justice*, Rawls brings up the distinction between ideal and non-ideal theory in the context of discussions about liberty. He initially refers to the two parts of non-ideal theory as "two kinds of circumstances that justify or excuse a restriction of liberty" (Rawls, *Theory* 215). This makes it seem as if non-ideal theory is only meant to address situations in which our liberty ought to be restricted. But it's not clear whether Rawls thinks that restrictions on liberty are definitive of non-ideal theory (they don't come up in his later discussions, such as in *The Law of Peoples*). We can make use of non-ideal theory without confining its application to restrictions on liberty, as Rawls himself does later.

Natural limitations and historical and social contingencies

That Rawls comes to non-ideal theory via considering restrictions on liberty affects the way he presents 2a2), natural limitations and historical and social contingencies. In Rawls's initial presentation of this branch of non-ideal theory in *A Theory of Justice,* it seems as though 2a2) is present in ideal as well as non-ideal conditions. Rawls writes, "even in a well-ordered society under favorable circumstances, liberty of thought and conscience is subject to reasonable

regulations…other [restrictions on liberty] are adjustments to the natural features of the human situation, as with the lesser liberty of children" (*Theory* 215). While Rawls intends natural limitations and historical contingencies to be part of non-ideal theory, it seems clear that some natural limitations and historical contingencies requiring restrictions on liberty (such as restrictions on the liberty of children) are present in the ideal world and thus belong to ideal theory.[4] Thus, we are left with a puzzle. It seems like some features of the ideal world are the subject of non-ideal theory. How can it be the case that non-ideal theory, which is theory about the non-ideal world, also governs features of the ideal world as well?

Perhaps remediability marks the distinction. There are some natural limitations that we could never remedy in any world we lived in. We will always have gaps in our knowledge and economic inequalities, maybe. Since these will persist in every world, they will persist in the ideal world. (We might notice that many of these limitations, such as the existence of children, don't even seem that bad.) It's only those limitations and contingencies that could be remedied—that don't exist in every world—that are part of non-ideal theory. Not all restrictions on liberty, that is, are non-ideal.

But I would suggest one small change to this formulation. Some natural limitations and historical contingencies might be path-dependent: that is, once we go down a certain path, our choices might be limited, including our choices for how to remedy a certain limitation or contingency. Once we've made some choices, a limitation or contingency might become irremediable. Once we mine all of our gold, it

[4] See also Thomas (12).

will be impossible to reach an ideal that involves making a lot of jewelry. But this kind of irremediability doesn't seem like it should affect whether that limitation is part of ideal theory. These problems wouldn't exist in the ideal world, even if we now have no way to get to that ideal. Thus, the distinction should be not whether a given limitation or contingency could *now* be remedied but whether it would be remediable *in some set of circumstances.* Only natural limitations and historical contingencies that are remediable *in no set of circumstances* are part of ideal theory.[5]

Once we have an idea of which natural limitations and historical and social contingencies are remediable, we will want to know what our duties of transition to the ideal are. What are we obligated to do with respect to making our world more ideal? Here, Rawls is not entirely clear. He writes that "as far as circumstances permit, we have a natural duty to remove any injustices, beginning with the most grievous…" (Rawls, *Theory* 216). But while 2b) deliberate injustice is (obviously) injustice, 2a) unfavorable conditions is not. Do we have a duty to remedy unfortunate circumstances that are nevertheless not injustices? Rawls isn't clear here, but maybe the duty extends to 2a) as well. We have duties to remedy non-ideal features of our society where we can, even if no one is to be blamed for causing those non-ideal features in the first place. If we fail to do this, we are doing injustice or at least allowing it to happen.

[5] An additional complication here is that Rawls writes that ideal theory "…develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the *fixed constraints* of human life" (*Theory* 216, emphasis his). Rawls appears to make a distinction between "fixed constraints of human life," which are a part of ideal theory, and "natural limitations," which are not. But if, as I have suggested, some natural limitations should be thought of as part of ideal theory, then it's not clear whether there's any meaningful difference between fixed constraints and (ideal-theory) natural limitations.

Casting the "natural duty" in this way preserves the intuition that we act unjustly when we don't do anything about the non-ideal features of our society, even if those non-ideal features weren't initially injustices.

Thus, the way to solve this puzzle is to admit that some natural limitations and accidents of human life can figure in and constrain ideal theory, while some don't. The place to draw the distinction between ideal and non-ideal limitations is whether it would ever have been possible for these limitations to be remedied. Limitations that can never be remedied—that are a permanent and necessary part of human life—are part of ideal theory, since the ideal world would necessarily have these limitations as well. In arguing for this, I am disagreeing with Simmons, who retains all of the natural limitations Rawls names as part of ideal theory, because they "all seem to involve departures from ideal principles" (16). And laws restricting the liberty of children don't seem to be necessarily pernicious features of the ideal—Rawls even allows for penal sanctions within ideal theory (*Theory* 212). Admitting that some natural limitations are part of ideal theory and drawing the distinction between ideal and non-ideal in terms of remediability is thus an amendment to Rawls, but (I think) a friendly one that preserves the intuitive force of the distinction he draws.

The relationship between ideal and non-ideal theory

For Rawls, "the ideal part presents a conception of a just society that we are to achieve if we can" (*Theory* 216). Ideal theory gives us guidance: it shows us which elements of the ideal are relatively more urgent, which helps to tell us how to derive non-ideal from ideal theory (Rawls, *Theory* 216). So non-ideal theory depends on

ideal theory.[6] In the non-ideal case, we judge whether institutions are just by whether they depart from the ideal without sufficient reason (Rawls, *Theory of Justice* 216). Our non-ideal institutions don't perfectly mimic ideal institutions, because they adapt to non-ideal circumstances. Rawls writes that "the measure of departures from the ideal is left importantly to intuition," and the measure to which our priorities in non-ideal cases should match our priorities in ideal cases may be left up to intuition as well (*Theory* 216). Because of this, ideal theory certainly isn't sufficient for guiding us in non-ideal situations.

When Rawls talks about our duties in non-ideal situations, the only duty he mentions is the duty to remove existing injustices (*Theory* 216). This suggests that he's more concerned with A) transitional non-ideal theory than with B) non-transitional non-ideal theory. But surely both have a place in non-ideal theory. If, as I discussed above, some natural limitations and historical and social contingencies are non-ideal and also not remediable (perhaps because of certain path-dependencies), then there are non-ideal situations for which we may not be responsible and which may have no solutions. Surely, then, we have non-ideal duties that are not also transitional duties. They may be different from our ideal duties, since they may be constrained by some limitations of our nature, history, or society, but they won't be focused on removing those limitations. Rawls's discussion of the duty to remove injustices doesn't exclude the possibility that we may have duties in the face of persisting injustice; we'll need both A) and B) non-ideal theory.

---

[6] See Simmons (10).

One place in *A Theory of Justice* that Rawls hints at how to develop B) non-transitional non-ideal theory is in his discussion of the general conception of justice. He describes the general conception this way: "All social values—liberty and opportunity, income and wealth, and the social bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage" (Rawls, *Theory* 54). This general conception is much more permissive than the special conception (the two principles of justice). The special conception requires that each person have "an equal right to the most extensive total system of equal basic liberties compatible with a similar system of law," but the general conception only requires that an unequal system be to everyone's benefit (Rawls, *Theory* 266). So, for example, in a society governed by the general conception of justice, we might all forgo certain political rights, if that will make us economically better off; we can't do that when we're governed by the special conception, because the first principle has lexical priority over the second (Rawls, *Theory* 55).

Rawls gives a couple of directions for how to institute the general conception: "If possible, the more central [freedoms] should be realized first," and we must make sure that we can bring about social conditions "under which restrictions on these freedoms are no longer justified" (*Theory* 216; 217). Thus, Rawls gives us a (very general) priority rule, and he also reminds us that the goal is to bring about conditions in which we can be governed by the special conception. If our problem is that our economy is too small to support basic liberties for everyone, we should not institute a

form of the general conception that prevents the economy from growing to the point where we can institute the special conception.

The general conception, then, might provide B) to go along with Rawls's A); while most of Rawls's non-ideal theory is directed at getting us toward the ideal, the general conception can tell us what duties we have to each other while in non-ideal circumstances. Simmons thinks of the general conception "as a kind of intermediate conception, lying between" ideal and non-ideal theory, but I don't think this is quite right (14). The idea of a continuum with ideal theory, the general conception, and non-ideal theory occupying different points suggests that non-ideal theory and the general conception are rival conceptions of justice and are instituted serially as societies become more or less just. But this doesn't appear to be what Rawls has in mind. Instead, we institute both transitional non-ideal theory and the general conception at the same time. The general conception tells us about justice, but it's not intended to be a permanent replacement for ideal justice; it gives us duties of justice *while* we transition to a state in which we can implement full ideal justice. The general conception works hand-in-hand with A) Rawls's transitional non-ideal theory. It's the companion B) non-ideal non-transitional theory to A) Rawls's transitional theory; it doesn't govern a different type of situation.

The distinction in Rawls's later work

Although *A Theory of Justice* and *The Law of Peoples* are where Rawls does most of his work in ideal and non-ideal theory, the distinction is at least mentioned in *Justice as Fairness: A Restatement* and *Political Liberalism* as well. In the former,

Rawls writes of ideal theory that "We ask in effect what a perfectly just, or nearly just, constitutional regime might be like, and whether it may come about and be made stable under the circumstances of justice, and so under realistic, though reasonably favorable, conditions" (*Justice as Fairness* 13). Thus, realistic but *reasonably favorable* conditions are part of ideal theory; presumably realistic and *unfavorable* conditions are part of non-ideal theory. This reinforces our answer to the puzzle about 2a), the natural limitations and historical and social contingencies. In parts of *A Theory of Justice,* Rawls seems to include in non-ideal theory even features ideal societies would have, such as regulation of speech and limitations on majority rule. But these seem like realistic but reasonably favorable conditions, and so they should be a part of ideal theory according to *Justice as Fairness*-era Rawls.

This is consistent with what Rawls writes in *The Law of Peoples,* his work on global justice. The taxonomy of societies Rawls offers reflects his ideal/non-ideal distinction. The subject of ideal theory is with "reasonable liberal" (often "liberal democratic") peoples and "decent" peoples (which aren't liberal democracies, but which obey the Law of Peoples and so count as members of the Society of Peoples) (Rawls, *Law* 4). These are what Rawls calls "well-ordered peoples" (analogous to the "well-ordered society" that is the subject of ideal theory in *A Theory of Justice*). Rawls lays out conditions for ideal global justice in *The Law of Peoples,* and these two kinds of societies subscribe to these conditions.

On the other side of the distinction, we have non-ideal theory and the societies that correspond to it. In *The Law of Peoples,* Rawls gives a slightly different account

of the two kinds of non-ideal theory than he does in *A Theory of Justice*. Here, "one kind deals with conditions of noncompliance," in this case refusal to comply with the Law of Peoples (Rawls, *Law* 5). It's unclear what to make of Rawls's claim that this kind deals with *non*compliance rather than *partial* compliance, as Rawls refers to it in *A Theory of Justice* (*Law* 8). Perhaps partial compliance is equivalent to noncompliance in the sense that *any* failure to comply with the Law of Peoples (or the two principles, in the case of domestic society) is noncompliance, whether that failure is partial or full. Or perhaps the shift in terminology reflects a shift in Rawls's thinking. [7]

Noncompliant societies are called "outlaw states." These states "refuse to comply with a reasonable Law of Peoples" (Rawls, *Law* 90). Given Rawls's language here, and the contrast with burdened societies that I discuss below, it seems that outlaw states *could,* but don't, comply with the Law of Peoples. So this first part of non-ideal theory is basically the same as 2b) in our original schema—the outlaw states are doing injustice.

So far, this reflects Rawls's thinking in *Theory of Justice* pretty closely. The most interesting contribution *Law of Peoples* makes comes in the discussion of the second kind of non-ideal theory (corresponding to the first part in *Theory of Justice*). This is the kind Rawls says "deals with unfavorable conditions, that is, with the conditions of societies whose historical, social, and economic circumstances make their achieving a well-ordered regime, whether liberal or decent, difficult if not

[7] It's unclear why Rawls shifts from *partial* compliance in *Theory of Justice* to *non*compliance in *Law of Peoples,* but it doesn't appear to reflect a substantive shift in his thinking.

impossible" (*Law of Peoples* 5). These are the "burdened" societies (*Law of Peoples* 5). These burdened societies are thus analogous to 2a), natural limitations and historical and social contingencies. Remember that above, we discussed the difficulty of pinning down what exactly counts as 2a1), natural limitations. Some of these limitations, as Rawls originally defines them, are present even in well-ordered societies (for example, reasonable regulations on liberty of conscience) (*Theory* 215).

But we saw that we might just take Rawls to mean that non-ideal natural limitations count as 2a1). If we go with that reading of Rawls, *The Law of Peoples* is very similar. The unfavorable conditions in *The Law of Peoples* are clearly not natural parts of the life of a society. Only some societies are constrained by historical, social, or economic circumstances that make it difficult or impossible for them to achieve justice. These constraints are avoidable, nonnecessary parts of human (or societal) life. So we avoid having to make the distinction between natural limitations and fixed constraints. And unfavorable conditions are much more clearly a natural part of non-ideal theory. Ideal societies are not burdened by unfavorable conditions; only non-ideal societies are prevented by these conditions from achieving the ideal.

When we put this together with the other kind of non-ideal theory, we get a clear blameworthiness/responsibility/control distinction, along the lines of the distinction I articulated before in connection with *A Theory of Justice.* Both burdened societies and outlaw states fail to comply with the requirements of justice; the distinction is that outlaw states *could* but do not, while burdened societies *cannot* and do not. Outlaw states can control/are responsible for their failure to be just, and are

thus blameworthy for this failure; burdened societies cannot control/are not responsible for this failure, and are thus not blameworthy for it.[8]

While *The Law of Peoples* is in general a clear account of Rawls's distinction, one puzzle remains. In *A Theory of Justice,* recall, 2a1) natural limitations were a confusing category of non-ideal theory, because at least some of them are present in the ideal as well as in non-ideal societies. So not all natural limitations seem like the subject of non-ideal theory. But perhaps some are—even if they're natural, perhaps we must remedy them in order to transition to the ideal. So while Rawls's treatment of natural limitations in *A Theory of Justice* is complicated, there seems to be a place for at least some natural limitations in non-ideal theory. But natural limitations aren't discussed in *The Law of Peoples.* The examples Rawls lists of the burdens of burdened societies—a lack of political or cultural traditions, human capital, and resources—seem to be 2a2) historical and social contingencies rather than natural limitations (*Law* 106). In *The Law of Peoples*, then, only these contingencies (as well as deliberate injustice) seem to count as non-ideal theory. What happened to natural limitations?

Here, I think we should retain the insight from our discussion of *A Theory of Justice.* In *The Law of Peoples,* Rawls focuses on contingencies in non-ideal theory. This may be because it's more obvious that we can correct these contingencies, since only certain individuals and societies suffer from them. But there may be an important place in non-ideal theory for natural limitations as well, since at least some natural

[8] Simmons comes up with something similar (16). Note as well the possibility—and Rawls acknowledges this—that states could be both burdened and outlaw. They could be incapable of complying with some requirements of justice and also fail to comply with those requirements of justice that they're capable of complying with. And surely the same is true of non-ideal theory in the domestic case

limitations may be remediable. I suggest the best way to think about this part of non-ideal theory is to frame it in the broader terms of blameworthiness, responsibility, and control. Societies, and individuals within those societies, are surely not to be blamed for having certain natural limitations, but we might still think that those limitations are the proper subject of non-ideal theory where they can be remedied. This fits with the way Rawls is thinking about burdened societies in *The Law of Peoples,* even if he doesn't discuss natural limitations explicitly: burdened societies have had burdens placed upon them, whether those burdens are natural or contingent features.

Considering both *A Theory of Justice* and *The Law of Peoples* gives us the clearest picture of Rawls's thinking on the distinction between ideal and non-ideal theory. The ideal is the goal we are supposed to pursue. It may contain some natural limitations, such as those on the rights of children, but those are only the limitations that are necessary features of human life. Non-ideal theory tells us what to do in the face of two different kinds of problems. Some are problems for which we are blameworthy; those are issues of deliberate injustice. Some are problems for which we are not blameworthy; those are remediable natural limitations and historical and social contingencies. Our goal, whether as members of a society or as members of the Society of Peoples, is to transition to the ideal and to help other societies make that transition as well.

1.2 Sen's critique of ideal theory

Amartya Sen, among others, thinks ideal theory is not helpful. In his critique of ideal theory in *The Idea of Justice,* Sen offers three criticisms of ideal theory—that it

is not sufficient to give complete guidance in non-ideal settings, that it's not necessary to give guidance, and that "totalist," or complete, ideal theories are likely to be impossible. In what follows, I reject the first two criticisms but accept some form of the third. Addressing Sen's criticisms of ideal theory helps us to see how ideal theory can provide guidance. We must develop ideal theory that both accomplishes Rawls's goals and recognizes Sen's important points about our theoretical limitations.

Sen's criticisms seem aimed squarely at ideal theories such as Rawls's, but he barely uses the term "ideal theory." Instead, he criticizes two overlapping approaches to theories of justice—the "transcendental approach" and the "totalist approach." The transcendental approach seeks to identify perfect justice, while the totalist approach searches for completeness in a theory of justice (Sen 11; 103). Sen says that these have usually gone together: the attempt to identify perfect justice has usually been accompanied by the attempt to come up with a complete theory of justice. Rawls's approach is transcendental, says Sen, because it identifies a single ideal of justice and totalist because it gives (or, at least, aims to give) a complete picture of what that ideal justice looks like.[9] In contrast, the approach Sen favors is comparative, rather than transcendental, and it allows for incompleteness, rather than being totalist. Sen prefers an approach that compares existing society with other possibilities that we can reasonably expect we could attain, realizes conflicts in principles and values may leave us with an incomplete theory of justice, and comes up with a partial ranking of possible societies based on the views of the people involved (106-11; 95).

[9] Sen calls Rawls's theory of justice transcendental at p. 95-6 and totalist at p. 103 of *The Idea of Justice*.

The transcendental approach

We start, as Sen does, with his critique of the transcendental approach. First, note the unusual terminology: the term "transcendental approach" is not commonly used by other philosophers writing on this subject. Sen says that this approach "tries only to identify social characteristics that cannot be transcended in terms of justice," rather than also comparing feasible but non-ideal societies (6). So it seems that the transcendental approach is hunting for social characteristics, or perhaps a society, that is so just that we could not think of a more just society or more just social characteristics. This seems straightforwardly like what other philosophers have called "ideal theory."

It's unclear why Sen uses "transcendental" rather than "ideal." On the face of it, it doesn't seem to have much to do with other ways "transcendental" is used in philosophy, such as Kant's transcendental idealism or his transcendental arguments.[10] A transcendental argument is standardly taken to be an argument that shows that, given uncontroversial premise(s) *X*, substantive conclusion *Y* necessarily follows (Pereboom; Stern). It's not clear from what premises to what conclusion Sen thinks transcendental theorists are reasoning.

To Alan Thomas, however, Sen's use of "transcendental" makes more sense.[11] Thomas's explanation for Sen's use of "transcendental institutionalism" is that the view is "*transcendental* in the sense that it appeals to a set of perfect principles of

[10] See the Stanford Encyclopedia of Philosophy articles by Stern ("Transcendental Arguments") and Pereboom and the section on Transcendental Idealism in the article by Rohlf for a primer on these issues.

[11] For a different view of what Sen might mean by "transcendental," see Valentini, "Paradigm Shift" (5-7).

justice (where 'perfect' here also has a precise sense: such principles offer a complete and transitive ranking of all possible social outcomes)" (2).[12] Thomas suggests that Sen actually *is* borrowing this term from Kant. Kant distinguishes between transcendence and the transcendental, where "*transcendence* is an appeal to unconditioned ideals that can have no application in our world," and the idea behind the *transcendental* "is that a certain style of argument can offer an immanent critique of our existing moral and political views while also giving us the critical purchase to go beyond them" (2). Perhaps Sen does mean to invoke Kant, but his argument is actually aimed at the kind of *transcendent* theory he believes Rawls has created. This makes sense: Sen is criticizing theories that claim to identify the most perfect social arrangement, the one that would be at the top of a ranked list. And when we use "transcendental" in this way, Sen surely has a transcendental theory as well, since he also offers critiques of our existing views.

The question then is whether this is an apt critique of Rawlsian ideal theory. Thomas suggests that it is not: "Sen has not demonstrated that Rawls believes in a 'perfectly just society' in the sense of that social outcome *than which no other can be ranked as more just*'" (10). Instead, Thomas thinks that Rawls's aim is to find the most reasonable conception of justice; this conception is not claimed to be perfect and unimprovable (10). If Thomas is right, then Rawls doesn't have a transcendent theory of justice, and so Sen's critiques miss their target. My aims are broader; as I discuss below, I think some of Sen's arguments against transcendental theories of justice fail

[12] All emphases here and in subsequent references are his.

even when directed against truly transcendental theories. But if Thomas is right, then Rawlsians have even more reason to feel secure in their views.

To Sen, this transcendental (or transcendent) approach doesn't give us the guidance we really need. It may tell us what some imaginary just world looks like, but this information is neither necessary nor sufficient to tell us how to remove specific injustices, such as hunger, poverty, and sexism (96). When Sen talks about the kind of guidance ideal theory could give us, he seems to have in mind the question of whether ideal theory is necessary or sufficient for ranking different possible societies. For example, when he first discusses whether the transcendental approach is sufficient, he asks, "Is the specification of an entirely just society sufficient to give us rankings of departures from justness in terms of comparative distances from perfection, so that a transcendental identification might *inter alia* entail comparative gradings?" (Sen 98). That is, for a perfectly just society to be sufficient in the way Sen conceives of sufficiency, it would have to be able, on its own, to tell us everything we need to know in order to rank all other societies in terms of justice. Similarly, Sen's conception of necessity is that the perfectly just society would be necessary for our being able to rank societies.

This is a difference between Sen and ideal theorists. For most ideal theorists, the ability to rank societies will take a backseat to the ability to transition from the current society to the ideal. The necessity and sufficiency these ideal theorists are concerned with, then, is whether knowledge of the ideal is necessary or sufficient for improving our society and, eventually, for reaching the ideal. Perhaps rankings of all

possibilities can help to give us information about how to reach the ideal, but ideal theorists have not usually emphasized rankings of possibilities as an important component of ideal theory. In order to consider Sen's critiques as they pertain to ideal theory, I will consider Sen's and ideal theorists' meanings of these terms in what follows.

Sufficiency

Sen starts with sufficiency. We already know that Rawls hasn't given a theory on which transcendental justice is sufficient to rank all alternatives; Rawls himself admits that judging the justice of non-ideal societies is partly a matter of intuition (*Theory* 216). But Sen wants to argue that it isn't even *possible* for the transcendental ideal to be sufficient to rank the others. He argues that the transcendental ideal cannot be sufficient to rank the others, because he doesn't see a way to judge which departures from the ideal are worse than others (Sen 98-101). For example, he doesn't see how to judge whether a violation of freedom of speech is worse than a violation of freedom of religion.

And ideal theorists agree. The only way an ideal would be sufficient would be if we knew the various weights of all different features of a society, but neither Sen nor Rawls thinks we have that kind of knowledge. Ideal theorists don't just want a theory of justice to rank all alternatives. We also want to know how to make progress toward the ideal where this is possible, and, where it's not possible, we want to know how we should behave. Knowing the transcendental ideal, even knowing the weights behind the various features of the transcendental ideal, doesn't seem like it will give us

a complete non-ideal or transitional theory. As Robeyns says, ideal theory may be able to tell us about the Paradise Island we would like to get to, but just knowing its location cannot tell us everything about how to get there (361). We need theory about whether our obligations in a non-ideal situation are different than in an ideal situation (part of non-ideal theory) and theory about what kinds of progress are permissible and which are impermissible (part of transitional theory). So ideal theory cannot be sufficient. Sen's critique succeeds as far as it goes, which isn't very far.

Necessity

So the transcendental approach is not sufficient to perform all the tasks ideal and non-ideal theory should perform. But even if the transcendental ideal isn't sufficient, can it still be necessary for guidance? Sen considers this question next. He points out that, in general, this would be an unusual thing to expect of transcendental ideals: "relative assessment of two alternatives tends in general to be a matter between them, without there being the necessity to beseech the help of a third—'irrelevant'—alternative" (Sen 101). We don't need to know what the most perfect painting looks like in order to judge that one painting is more beautiful than another.

Remember that the necessity Sen's talking about here is distinct from the necessity ideal theorists are generally concerned with. What ideal theorists are generally concerned with is whether the ideal is necessary for guiding our transition from the non-ideal world we live in to the ideal, or at least to a better world. Knowing how two non-ideal worlds rank relative to each other may or may not be important for this end. If choosing the better of two non-ideal worlds will actually get us further

away from the ideal, then knowledge of how they rank relative to each other won't be very helpful. A society with very high taxes might seem further from the ideal than a society with very low taxes, but we might need the high taxes in order to fund critical defense and infrastructure projects.

In order to make the argument that it's not necessary to know about transcendental justice, Sen gives an analogy: "we may indeed be willing to accept, with great certainty, that Mount Everest is the tallest mountain in the world, completely unbeatable in terms of stature by any other peak, but that understanding is neither needed, nor particularly helpful, in comparing the peak heights of, say, Mount Kilimanjaro and Mount McKinley" (102). We don't need to know the height of the tallest mountain in order to measure other things; similarly, we don't need to know what the most just society is in order to determine which of two societies is more just. Knowing the transcendental ideal form of a given thing, whether it's the height of a mountain or the justice of a society, will not give us guidance on which of two alternatives is higher or more just. So the transcendental ideal isn't necessary.

There are two reasons this analogy is mistaken, and these reasons help us to see why a transcendental ideal can be necessary for the purposes ideal theorists have in mind. The first is that Mount Everest is not a transcendental ideal. Sen calls it "completely unbeatable in terms of stature by any other peak," but that's not right. Mount Everest is instead completely unbeatable in terms of stature by any other *existing* peak. Maybe you could hold that knowing about the most just existing society allows you to rank all alternatives, but this is not the view of ideal theorists; they're

looking for the *ideal* society, not the *best existing* society, because only the ideal society instantiates perfect justice.

So the analogy seems inapt. But we might still question whether the tallest possible mountain (call it Mount *X*) is any help to us in comparing the heights of other mountains. What can Mount *X* tell us about the relative heights of Mount McKinley and Mount Kilimanjaro? Not much. But the reason that neither Mount Everest nor Mount *X* is necessary for figuring out whether Mount McKinley is higher than Mount Kilimanjaro is that we have a criterion for measuring mountains that's independent of the highest mountain—that is, the criterion of height. We can figure out the height of Mount McKinley in units (feet or meters) that are completely independent of our knowledge about Mount *X*.

That's not true for justice. Unless and until we flesh out justice in a particular way, (which I discuss below), we don't have a criterion for measuring justice. We don't have theory-neutral units of justice that we can use to measure societies in order to say that Society *P* is more just than Society *Q*. Because we don't have an independent criterion of justice, ideal theorists look to the transcendental ideal to give us some basis for comparison. Consider a case that's more closely analogous to justice. The situation we're in with respect to justice is like if we, as non-geographers, took a trip to the Himalayas and tried to figure out which mountain is which. We might know that Mount Everest is taller than K2, and so on, but the way we can figure out which mountain is Mount Everest and which is K2 is to compare the highest mountain we can see to the others. Without a criterion of height, we must compare

mountains to each other, and we can start by identifying the tallest mountain. Here, the different mountains give us the criterion by which we compare mountain heights, just as in the case of justice the transcendental ideal gives us the criterion by which we compare societies to each other.

The exception to this line of argument is the group of maximizing views about justice, such as the views that justice consists in maximizing general well-being or rights. Maximizing views cannot have a transcendental ideal, because more of the good that's being maximized is always better. There's no ideal we can reach and so stop maximizing the relevant good. And these views have a way to compare societies without reference to an ideal—the society with more of the relevant good is the preferable society. Just like we have the independent criterion of height to compare mountains, these views have independent criteria of justice that they can use to compare societies. So if one of these views turns out to be right, we won't need a transcendental ideal to compare societies.

Notice, though, that Sen isn't making an argument here for one of these maximizing views—he's making an argument that the transcendental ideal is *never* necessary for guidance.[13] People who find maximizing views plausible for other reasons may find Sen's arguments against the necessity of the transcendental ideal plausible, just as people who don't find maximizing views plausible—people who don't think there's a criterion of justice analogous to height—may not. So Sen's arguments aren't independently convincing. They certainly can't convince someone

[13] See also Valentini, who writes that "Unless Sen is prepared to deny this substantive claim [that justice has a cutoff point beyond which it can no longer be maximized], he cannot dismiss the value of theorizing about perfect justice so easily" ("Paradigm Shift" 8).

like Rawls, who holds (independently of arguments about ideal and non-ideal theory) a non-maximizing view of justice. And we have independent reason to think the ideal is necessary.

A positive argument for the necessity of the ideal

That independent reason is this. Critics of ideal theory often use the problem of second best as evidence for their view. This theory, which comes out of economics, shows us the difficulty of simply reading the second-best outcome off the features of the best outcome.[14] What makes something the best isn't necessarily just the sum of its features; these features may also interact with each other. This means that the second-best outcome will not necessarily be the one that most closely resembles the best outcome. For a simple example, consider making dinner. If I have pasta, then I'll prefer to make a dinner of pasta, marinara sauce, and salad. But if I don't have pasta, then I don't want marinara sauce either—my preferred dinner would be a burrito, rice, and beans, rather than pasta, marinara sauce, and salad. That is, I would prefer the non-ideal outcome that has *fewer* (or none) of the features of the ideal to the one that has *more* of the features of the ideal. Critics of ideal theory argue that since we can't read off what we should do simply by what's most similar to the ideal, we can't trust the ideal to give us guidance.[15]

And yet theories without ideals also suffer from a version of this problem. If we don't have an ideal guiding our progress over time, all we have to go on is comparative judgments about which states of affairs are better than others (Simmons

[14] For the original paper in economics, see Lipsey and Lancaster. For a sample explanation in the philosophical literature, see Goodin.
[15] For example, see Wiens, "Prescribing Institutions."

23). We may be able to make a decision about which of two states of affairs we prefer, but this isn't a recipe for sustained progress. We might be tricked by attractive features of one state of affairs, even though choosing that state of affairs will not be beneficial in the long run. To keep going with the dinner analogy, I might first be faced with the choice between pasta and a burrito, and I might start cooking the pasta. But then I discover that I'm out of marinara sauce. I should have considered the interactions between the features of the possible alternatives before I started in on making my comparative judgments between dinners. Ideal theory may be necessary because we need to know what we hope to achieve before we start taking steps toward it.

Notice that this is true in the case of maximizing views as well. We might opt for the policy that maximizes a certain good in the short run, only to find that it doesn't maximize that good in the long run. Take a view on which justice consists in maximizing rights. Granting full property rights before we have the right redistributive mechanisms in place may lead to stunted speech rights later on. If we don't have some kind of ideal to guide our progress, we're vulnerable to the problem of the second best: we may only have comparative judgments to go on, and they may lead us astray. This isn't in itself an argument against maximizing views, but it is a suggestion that proponents of these views may need to consider accepting some kind of practical ideal for policy purposes; in this case, the ideal might be the greatest amount of the good to be maximized (rights or whatever) that is feasible, and states of affairs can then be judged relative to how well they bring about that amount of the good. (Then, of course, as circumstances change, we can raise or lower the amount of the good that

serves as the "ideal.") Thus, even maximizing views may find a sort of ideal practically useful, if not strictly theoretically necessary.

Because the ideal is necessary for keeping us from getting sidetracked by misleading pairwise comparisons between states of affairs, it's helpful to see why Sen's other analogy fails too. Sen points out that we do not need to know what the best painting is like in order to know whether a Van Gogh or a Picasso is better (101). But the case of ideal theory is like if we're trying to use the best possible painting in order to try to improve our own art. Here it is helpful to consider the features of this painting that makes it the best. If the best painting has blue in it, that doesn't mean a solid blue canvas is great art: we must consider the ways its other features interact with its use of blue. This improvement in our own art does not require us to produce a complete ranking of all art. In the same way, making progress over time does not require a complete ranking of alternative states of affairs, but it does require consideration of how the features of the ideal interact and which states of affairs will allow us to transition to the ideal. A good ideal theory gives us this information.

To turn to an example from political philosophy, consider race-based affirmative action. Grant for the sake of argument that affirmative action in college admissions is effective in moving us toward a society in which the races are equal. Establishing affirmative-action programs might seem on its face to be a case of setting back the cause of equality between the races, because it gives one race some sort of advantage over another in college admissions. Thus if we make a pairwise comparison between otherwise similar societies, one with affirmative action and one without, the

one without affirmative action seems to have more racial equality, and if we were choosing between societies on the basis of racial equality, we'd pick that one. From some perspectives, then, choosing not to institute affirmative action appears to be an improvement. Over time, however, the society with the affirmative-action program better achieves full racial equality than the society without it. The society that more closely resembles the ideal, that is, is not the society that gets us closer to the ideal as we make progress over time. If we're just making piecemeal progress by hopping from one comparative judgment to the next without a clear idea of where we want to end up, we fall prey to this problem.

Thus the problem of second best turns out to be a problem for critics of ideal theory. Where we might have thought that the problem of second best shows that knowledge of the ideal leads us astray, in fact we wind up with an argument for ideal theory. Without knowing what we're aiming for, a simple ranking of two states of affairs may not give us the kind of sustained progress we're after.[16] The problem of second best is a problem for *bad* ideal theory: ideal theory that doesn't take into account the interactions between features of the ideal and instead just makes progress by looking for whatever is superficially most similar to the ideal. It is not a problem for ideal theory done well. But it's a problem for *all* non-ideal theory done without an ideal. Since this kind of non-ideal theory doesn't have an integrated conception of

[16] Even if we hold a maximizing view, on which there's no one ideal because more justice is already better, we may need to adopt some kind of ideal for policy purposes in order to avoid being tricked by gains that seemingly maximize justice while actually slowing its progress. See the distinction between practical and theoretical ideals below.

where it will take us, progress must be piecemeal and therefore susceptible to the problem of second best.

Three kinds of necessity

Affirmative action is a particular, very complicated, case. While the ideal theorist wants ideal theory to provide guidance for continued progress over time, the critic of ideal theory can respond that it's frequently unnecessary. Jim Crow laws are obviously unjust; here, thinking about ideal theory doesn't give us additional guidance, because it's clear we should just do away with these laws. [17] Ideal theorists can respond to this challenge in three different ways.

Start with the least concessive response. The ideal theorist can hold his ground and insist that we need ideal theory no matter what. This is because having opinions about which policies are just or unjust is different from making claims about how those policies ought to change. We need to know what kind of society we're aiming for in order to know *how* to improve an unjust situation, even if we don't need this kind of ideal theory in order to know *that* the situation ought to be improved. In the case of Jim Crow laws, we need to know whether we should simply remove those laws, leaving a society that's formally equally under the law but in fact highly racially segregated, or whether we should remove those laws and replace them with new laws that ensure *de facto* as well as *de jure* equality. Only ideal theory, this response concludes, can tell us this.

---

[17] Valentini: "a society in which people are arbitrarily arrested is *obviously* more unjust than one in which, all other things equal, they are not. No account of perfect justice is needed to make this kind of judgment. Although correct, this observation is also rather inconsequential" ("Paradigm Shift" 8; see also "Conceptual Map" 661).

A more concessive response is to say that opponents of ideal theory are in fact doing ideal theory, just at a general level. In order for us to know that segregation is bad, we must have some idea of what justice and the ideal society look like (that is, that the ideally just society has racial equality). If these opponents of ideal theory did better ideal theory—if their ideas about the just society weren't inchoate and abstract—we would have more guidance across a broader range of cases.

Neither of these responses is likely to satisfy a critic of ideal theory. There may be some cases in which we truly don't need to know how to improve a policy; all we need to know is that the policy should go. Maybe the Jim Crow laws are like that: making the single improvement of striking down those laws is simply the right thing to do, no matter what replaces them. So the least concessive response doesn't work in all cases. Neither does the second response. An opponent of ideal theory might very well be doing inchoate ideal theory in some cases, but not in all. The opponent might be working off of values he thinks are the right values without considering what an ideal society that embodies those values might look like. So neither of the first two responses is sufficient to establish the necessity of ideal theory.

There's a third response, though, that is the most concessive but that covers the broadest range of cases.[18] This response starts by conceding the point about Jim Crow laws. Sometimes we don't need ideal theory. But we do need it when we've removed the Jim Crow laws and are dealing with less obvious and more complicated injustices. That is, we need ideal theory in order to make sustained societal progress over time.

[18] See also Stemplowska and Swift (8); and Robeyns (344-45).

Where the ideal is a big, complicated one (such as in the case of the ideal of the just society), and the problems keeping us from reaching the ideal are messy, and it will take significant time and effort to transition to the ideal, ideal theory becomes crucially important.

Think of it this way. If I want to get from the base of a mountain to the top, and I don't have a map, the first steps are relatively easy.[19] I look for a path, and I start hiking. If I happen to pick the wrong one—one that goes around the mountain instead of going up—I'll figure it out after I've hiked for a while, and I'll choose a new path. I may make mistakes in the first few steps, but usually the beginning is pretty self-explanatory. The tricky part is when the walking path peters out toward the top and I have to start climbing. Walking up the mountain was easy, but getting all the way to the top is much harder—without prior knowledge of the mountain, how do I know what climbing routes there are, let alone which are safe and which are dangerous?

The same thing is true of progress with respect to justice over time. The ideal theorist might have to concede that the first step—the abolition of Jim Crow laws—doesn't require ideal theory. What the ideal theorist does not have to concede is the necessity of ideal theory for guiding how to make the changes that come after Jim Crow—whether we should have laws aimed at bridging a gap between the races, what form those laws should take, and so on. We might mess up at the very beginning, just like I might take the wrong path at the start of my trip (and, perhaps, just like the Supreme Court did when it ruled in favor of "separate but equal" in *Plessy v.*

[19] Compare Simmons (35).

Ferguson), and ideal theory is helpful there too. But where it becomes necessary is after we've made the big obvious changes that are certain or likely to have few or no unforeseen negative consequences. We then have to make much finer-grained changes, and the potential for getting lost is much greater. This is where our map—ideal theory and associated non-ideal and transitional theory—becomes necessary if we want to avoid the problem of second best. As Valentini says, "Unless we want to content ourselves with our unsystematic and diverging intuitive judgments, Rawlsian-style higher-order moral reasoning becomes unavoidable" ("Paradigm Shift" 9).

The critic of ideal theory might say that this just shows that some kind of goal is necessary for guiding our progress, even if it's not the absolute ideal. With respect to the case of affirmative action, for example, we might think that the relevant goal is not the absolute ideal of complete equality between the races but instead a subsidiary goal of giving underserved minorities a fairer shot than they have now. We don't have to have the path to complete racial equality fully mapped out in order to be able to make real progress on an important goal. But the problem we saw above with pairwise comparisons repeats itself. We may make real progress toward the local peak, but we may get stuck there and be unable to make progress beyond it to the absolute peak. The better idea we have of where we want to end up, the better our chances of making real progress.

This doesn't yet tell us the extent to which the ideal can give us guidance. Perhaps we discover what the ideal is but don't think we should try to pursue it, because the costs of transition are too high. Perhaps we can't reach the ideal without

impermissibly trampling on the rights of others. The ideal is necessary for us to make sustained progress over time, but it's not sufficient—we also need non-ideal theory, to tell us what our duties are in the non-ideal world, and transitional theory, to tell us how to transition from non-ideal to ideal. These other theories can help us decide whether the ideal is something we should pursue. We should try to discover the ideal, because if it's attainable and we never find that out because we never look for it, we will not make the progress we could be making. Even if we should not try to reach the ideal, because the moral costs are too high, we may still find that we need it to guide our progress as far as we can go. As Rawls suggests, our judgments about non-ideal theory can be guided by our judgments about ideal theory (*Theory* 216). But ideal theory is necessary to ensure that we won't get stuck.

Completeness

So Sen's arguments against the transcendental approach have serious flaws. We saw that we have good reason to think that the transcendental ideal isn't sufficient to give guidance, but we don't have a reason yet to reject its being necessary—in fact, we need the transcendental ideal to provide guidance for sustained societal progress over time. But Sen has one more card to play. He argues that a theory of justice should not take what he calls a totalist form, although he says that "the standard theories of justice," including Rawls's, take this approach (Sen 103). By "totalist," Sen might mean one of two different things. First, totalist theories might be those that can give us a complete ranking of alternatives. Sen contrasts totalist theories with incomplete theories, which do not have "to find highly differentiated assessments of every

political and social arrangement in comparison with every other arrangement"—this sounds like he's thinking about how arrangements are to be ranked relative to each other (103). Second, Sen might mean that ideals are totalist when they give a complete picture of what the ideal world looks like. On this reading, a totalist ideal theory would tell us everything about the ideal: all laws, policies, principles, and other features of the ideal world. Either way, on this totalist approach to justice, "incompleteness tends to appear as a failure, or at least as a sign of the unfinished nature of the exercise" (Sen 103). Since we saw above that ideal theorists are generally more concerned with what the ideal is than with how to rank all states of affairs, I'll concentrate here on the second kind of totalism—on the idea that it's possible to get a complete picture of the ideal.The best way to understand the "totalism" of ideal theorists is as the claim that it is possible to arrive at a complete picture of the ideal.[20]

Against these "standard theories," Sen argues that incompleteness isn't a problem: "A theory of justice that makes systematic room for incompleteness can allow one to arrive at quite strong—and strongly relevant—judgments," such as that famines are unjust and that women are treated badly in many parts of the world (103). That's good, because totalist ideals may be impossible for several reasons, "including unbridgeable gaps in information, and judgmental unresolvability involving disparate considerations that cannot be entirely eliminated, even with full information" (Sen 103). These are reasons to expect incompleteness to be a permanent feature of our

[20] As with the transcendental approach, Sen could also mean the ability to come to a complete ranking of states of affairs, but since this isn't what ideal theorists are usually concerned with, I'll set that possibility aside.

theorizing about justice. Trying to do totalist ideal theory is a waste of our time, because this project is likely to fail.

What are the unbridgeable gaps in information that cause incompleteness in a theory of justice? Sen doesn't define them, but one example might be knowledge about the future. We may never be able to predict that a natural disaster will disrupt our ability to benefit the worst off. Sen's clearer on what judgmental unresolvability consists in. We might not be able to resolve whether it's permissible to sacrifice small gains in liberty for massive gains in economic equality (Sen 104). But these gaps in our knowledge don't stop us from wanting to do something about famine, or sexism, or extreme poverty, or torture.

A different kind of incompleteness arises when people who hold different theories of justice (whether they're complete or incomplete) try to come to an agreement. Sen writes that even after we remove partiality to ourselves using the veil of ignorance, "there may remain possibly conflicting views on social priorities, for example in weighing the claims of needs over entitlement to the fruits of one's labor" (104). If we have these different priorities, we cannot together come to an agreement on what justice looks like or on how to make a complete ranking of alternatives. But we can agree on some things—we can agree that a society that is neither just nor equal should go to the bottom of the ranking.

Sen writes that some incompleteness is assertive: that it yields "statements such as *x* and *y cannot* be ranked in terms of justice"; he contrasts this with tentative incompleteness, which holds only that we haven't yet found a way to reconcile *x* and *y*

(107).[21] We might question the existence of true assertive incompleteness, however: might it just be that we haven't yet figured out all the answers to ranking questions, even if those questions do have answers? Although Sen believes assertive incompleteness exists, he doesn't offer an in-principle argument for it, but we can at least agree that tentative incompleteness is pretty likely to hold in the conditions we find ourselves in. Even if it could be shown that elimination of incompleteness within a theory of justice is possible in principle, Sen might be right to think that the safe betting is on information gaps and judgmental unresolvability, at least for the time being. But we still need to make progress on justice, and we can't do so if we sit around and wait to see whether a complete ideal shows up.

So now we seem to be stuck. We saw reasons to think ideal theory is necessary when considering the possibility of sustained progress over time, and yet a complete ideal seems to be impossible, at least for now. The solution is to develop incomplete ideal theory. We must use the information we have to construct an ideal that's as complete as possible while at the same time accepting that a complete ideal may be out of our reach for now.

In his later work, Rawls himself denies that it is necessary to have a comprehensive conception of justice governing a society.[22] Because disagreements about justice are deep, we should instead come to an overlapping consensus: we should look for principles that all reasonable conceptions of justice can agree on and use those to govern our society. The same kind of consensus could form the basis of

[21] Emphasis his.
[22] See *Political Liberalism,* particularly Lecture I and Lecture IV.

an ideal theory. This ideal theory cannot be totalist, because it will prove to be impossible to reach agreement on all parts of the ideal, given our deep disagreements about what is best for our society. But it can be transcendental: it can show us what we should ultimately aim for. This kind of ideal theory does not fall prey to Sen's arguments against totalist theories of justice, but it preserves the structure of Rawlsian ideal theory and accords with the insight that we need ideal theory in order to ensure we can make sustained progress over time. So how can we do this?

1.3 Ideal theory after Sen

Recognition that we are in non-ideal conditions for doing ideal theory leads to a distinction between two types of ideal theory: the theoretical and the practical. Theoretical ideal theory is familiar; its object is to figure out what the transcendental ideal is. The practical kind is slightly different; its object is to figure out what ideal we can use to guide our progress. We can engage in theoretical ideal theory without ever expecting that this will produce usable results. A true-believer anarchist might realize that engaging in discussions about the finer points of anarchy is never going to push society in the direction of anarchy, but she might do it anyway, because she believes that the ideal is worth knowing about for its own sake. But when we do practical ideal theory, it is with an eye to figuring out how we can apply it in making decisions. This doesn't mean that the ideal must itself be feasible, because an infeasible ideal could guide our progress as well, by giving us something to approach. But the ideal must be able to guide our progress.

Making this distinction between practical and theoretical ideal theory concedes a lot to Sen. First, it concedes his point about the epistemology of our ideal theory—that gaps in the information we have will make it impossible for us to craft an immediately usable ideal theory. More importantly, however, it concedes that theoretical ideal theory does not necessarily guide our practice. Theoretical ideals may not gain the kind of popular support we need in order to make progress towards justice, because we may not all agree on which theoretical ideal to adopt. Rawls's two principles of justice are only useful ideal theory insofar as they can guide us to the best outcome possible for our society. This isn't to say that theoretical ideals don't have their place; as philosophers, we should be the last to think it's wrong to try to discover the truth for the truth's sake. And it may be possible that practical and theoretical ideals coincide, in the cases in which the complete truth about the ideal also guides our progress. But because of our limitations, we may not be able to use the theoretical ideal to guide our progress, and Sen is right to appreciate this limit to ideal theory. Ideal theorists, however, can stand our ground and insist that some kind of ideal is necessary for guiding our progress, as the arguments in the previous section of the chapter help to show. So while the practical/theoretical distinction makes significant concessions to Sen, it does so without conceding the need for some kind of ideal theory.

For the rest of this chapter, I'll limit my discussion to practical ideal theory. This limitation is friendly to Rawls, since his ideal theory is something we should try

to "achieve if we can," and it extends "the limits of practical political possibility" (*Theory* 216; *Law* 6). We want an ideal that can guide our progress.

Rawls and Sen constraints

From what I've suggested so far, we can come up with two constraints for our (practical) ideal theorizing: the Rawls constraint and the Sen constraint.

> **Rawls constraint:** The ideal is necessary for guiding our actions in order for us to make sustained societal progress over time.
>
> **Sen constraint:** Incompleteness within and across theories is a lasting feature of our theorizing.

Any plausible practical ideal theory must live within these constraints. We should not assume that we can proceed without identifying an ideal (we may need that ideal in order to make sustained progress over time), nor can we assume that we will be able to come up with a complete ideal.[23]

On the other hand, these constraints don't apply to theoretical ideals. Because their purpose isn't to guide our progress (and thus the Rawls constraint doesn't apply), we can keep working on them in hopes that we can find a totalist ideal (meaning that the Sen constraint won't apply). But if we want to get started now on transitioning toward justice, we will need to do practical ideal theory, and these constraints will apply. I devote the remainder of this chapter to trying to find a space for ideal theory within these two constraints. On my view, practical ideal theory should be reconceived

[23] It's possible that there are some cases in which the Rawls and/or Sen constraints fail to hold (that is, cases in which the ideal isn't necessary for us to make sustained progress over time, and/or cases in which we can come up with a complete theory), but we can't assume from the outset that either of these constraints will fail to hold.

as a process: we come up with agreements on what we want the ideal to look like, then we move toward that ideal, and then we readjust and refine our conception of the ideal as we transition toward it. This process won't guarantee that we transition to the *ideal* ideal, since it won't guarantee us a complete theoretical ideal theory, but it's the best way to make sustained progress over time given the Rawls and Sen constraints.

Political liberalism

Although I've named the Sen constraint after Sen, we should note that Rawls has already thought about a closely related problem. In *Political Liberalism,* he asks how political disagreement comes about between reasonable people. Rawls blames this on the burdens of judgment, which are "the many hazards involved in the correct (and conscientious) exercise of our powers of reason and judgment in the ordinary course of political life" (*Political Liberalism* 56). These burdens include the difficulty of assessing complex evidence, the vagueness of our concepts, and the way our different experiences shape our assessments (Rawls, *Political Liberalism* 56-7). If my life experiences cause me to weigh justice more heavily in my deliberations, and yours cause you to weigh benefits to the worst off more heavily, then we may be unable to reach an agreement about which policy to implement, even if we can both agree that justice and benefits to the worst off are both worth promoting. The burdens of judgment explain how everyone can be fully reasonable and still be unable to reach an agreement.

Because of the burdens of judgment, we are unable to come to a consensus on a single comprehensive doctrine, and thus the state may not coerce us to follow one,

even if there really is a single true comprehensive doctrine. Instead, Rawls develops the idea of an overlapping consensus of reasonable comprehensive doctrines: from the standpoints of our different comprehensive doctrines, we endorse a free-standing political conception of justice that may be supported by these comprehensive doctrines but does not depend on any one of them (*Political Liberalism* 134; 12-13). Instead, the political conception of justice we find in the overlapping consensus comes out of the fundamental ideas of the liberal political tradition (Rawls, *Political Liberalism* 14). The overlapping consensus is stable because everyone has his own reasons to endorse it coming out of his own comprehensive doctrine (Rawls, *Political Liberalism* 143).

Rawls doesn't spend much time on the ideal/non-ideal distinction in *Political Liberalism*, but this work has lessons for the project of constructing ideal theory. As we saw in making the distinction between theoretical and practical ideal theory, in an ideal epistemic situation we would be free of the burdens of judgment—we would be able to fully assess all the empirical evidence, our concepts would not be vague, we would be able to abstract from our personal experiences in weighing the evidence, and so on. We could perhaps reach agreement on a comprehensive doctrine. But we're not in that situation, because our cognitive powers are not infinite. Because of the burdens of judgment, we have reasonable pluralism in the non-ideal world—different people adhere to different reasonable comprehensive doctrines.

The congruence of this Rawlsian point with Sen's argument against completeness should be fairly clear. Where Sen has unbridgeable gaps in information and judgmental unresolvability, Rawls has the burdens of judgment. Even if we could

construct a complete comprehensive doctrine, the burdens of judgment would likely prevent everyone from accepting a single comprehensive doctrine. Thus, Rawls and Rawlsians (at least those who accept Rawls's turn to political liberalism) should recognize and accept Sen's argument against completeness in the case of ideal theory. Even if you and I can construct complete ideal theories, the burdens of judgment may make it impossible for us to reconcile our different ideals. This doesn't matter for the purposes of constructing a theoretical ideal, but it matters if we want to persuade others that our ideal is the right ideal, and it certainly matters if we want to agree on which policies to adopt.

Political Liberalism gives us a model for how to work around the cognitive problems that lead to this incompleteness—we look for principles of justice that are justifiable to all reasonable people, and we don't use one comprehensive doctrine as the basis for our political system. Ideal theory must work in a similar way. We must find ideals that reasonable people can accept, and we must accept that everyone will not agree on one theoretical ideal theory. As we will see, the process for coming up with an ideal theory diverges from the process for coming up with principles of justice.[24] But at a fundamental level, the two projects face the same question of how to make progress in a diverse, pluralistic society.

Practical ideal theory as a process

So when we do ideal theory, we are doing it in non-ideal circumstances. Our cognitive limitations will probably prevent us from coming up with a complete ideal

[24] And acceptance of the process of practical ideal theory, as I have outlined it, does not require accepting political liberalism.

theory that is acceptable to everyone from the outset; there certainly aren't any prospects for this at the moment. Rather than conceiving of ideal theory as something that we think up, complete, and then try to achieve, we should expect to develop and refine our ideal theory over time. Since we know that we may never come up with a complete practical ideal theory, we must simultaneously develop it and work to achieve it.

This is a process of coming up with particulars, crafting an ideal out of those particulars, figuring out how to transition toward that ideal, and then adjusting the ideal as our theories become more complete. The process requires us to come to a consensus about the particulars to include in the ideal; it also depends heavily on guidance from social science about how to make progress. This procedure resembles scientific research: scientists come up with hypotheses, test them, build theories that aim to give general explanations for phenomena, use these theories to do work in the real world, and then revise the theories as new information becomes available. Scientists must work with incomplete information, but they must also try to fill in the gaps in their knowledge. When we do practical ideal theory, we must do something similar.

Incompletely theorized agreements

In Rawlsian political liberalism, the aim is to come up with principles of justice that can attract the support of an overlapping consensus. Practical ideal theory can take a different approach, by starting by hunting for particular features of the ideal. When we think about ideal theory, we're thinking about what the ideally just

society is *like:* about the various features it might have. What would be the facts in the ideal world about how people are treated, what goods they have, and so on? The ideal gives us the goal that we're attempting to reach. We may be able to extract principles from this (from the particular "private property would be secure in the ideal" we may extract the principle "it is wrong to steal"), but starting with the particular features of the ideal world gives us a way to construct the goal we ought to be trying to reach. Thus, although Rawls's political liberalism provides us with evidence that Rawls anticipated the problems of inadequate and conflicting information and judgments, and although it provides us with some help for understanding how to reach agreement on ideal theory, the procedure for agreeing on principles of justice will be different from the procedure for constructing an ideal theory.

One way to do this is by using what Cass Sunstein calls incompletely theorized agreements. These are agreements "on the result and on relatively narrow or low-level explanations for it," on the fundamental principles behind a result (Sunstein 1735-6). For example, I might believe that murder is wrong because it's against God's law, and you might believe it's wrong because it destroys human dignity, but we can converge on an incompletely theorized agreement that murder is wrong. We can then apply this agreement in making real decisions: we can agree to punish a murderer even if we can't agree on the reasons for doing so. Sunstein writes that these agreements are "an important source of social stability and an important way for diverse people to demonstrate mutual respect," because people can respect that others have different but

reasoned sets of fundamental principles they use to decide cases (1736). Incompletely theorized agreements allow us to get things done in society without first agreeing on every principle and theory.

On Sunstein's method, incompletely theorized agreements make use of as little abstraction as possible. So, for example, we may be able to protect endangered species, even if we can't agree on whether this is because of the benefit to humans, or to the species themselves, or to the environment as a whole (Sunstein 1736). Protection of endangered species is a particular kind of program a society should undertake, and this program can be undertaken even when the rationales for it differ. In this case, for example, we can make a rule that we should protect endangered species, and we can come to an incompletely theorized agreement as to how to do it—by setting aside nature preserves, keeping endangered animals in zoos, and so on.

Sunstein mentions two different methods of reaching incompletely theorized agreements: rules and analogies (1743). We could be able to agree on the meanings of rules and that those rules are good, even if we can't agree on exactly why the rules are good. We'll be able to agree on how to apply those rules in many cases. For example, we can probably agree to establish a rule against committing murder, and we will find ourselves able to agree in most cases whether a killing is a murder or not. Analogies may help in more difficult cases. If it's not clear whether some particular killing counts as a murder, we can reason together about particular cases we agree on and see whether the tricky case has the relevant features of those cases.

Sunstein's reasoning is similar to Sen's: he recognizes that people may not have arrived at a complete theory that completely explains the result, or, if they have, they might not believe they can come to an agreement with others on it (1737). This echoes Sen's point that we should expect incompleteness both within a single theory of justice and when working with multiple theories of justice. Also like Sen, Sunstein recognizes that pluralism is a persistent feature (in his case, of our legal system), and yet we have to make progress: "Decisions must be made rapidly in the face of apparently intractable social disagreements on a wide range of first principles" (1735). Even when we have intractable disagreements at the level of our fundamental principles, we may be able to make progress on the particulars.

This isn't a foolproof technique; sometimes, the way that fundamental principles disagree with each other has ramifications for the policy choices we make. For example, it might turn out that some endangered animals are of no conceivable value to humans. If we think that we should protect endangered animals for their inherent value, we will protect these; if we think that we should protect endangered animals for their value to us, we will not. Yet often, Sunstein says, we can "reach a degree of closure by focusing on relative particulars" (1736). In the face of the burdens of judgment, this may be the best we can get.

One of Sunstein's justifications for incompletely theorized agreements is particularly relevant for our purposes: he writes that "incompletely theorized agreements may be valuable when what is sought is moral evolution over time" (1749). Completely theorized judgments lead to inflexible rules—since the judgment

has a complete theory behind it, what could be the reason for making an exception to the rule? But we know that we could be wrong about some of our values—Sunstein gives the example of equating homosexuality with incest—and so we want to be open to revisions (1749). This point is relevant to how to do ideal theory because of the incompleteness we expect to see across ideal theories (as well as within them). As we make theoretical and practical progress, we can expect that our ideal theory will become more complete, but we can also expect that it will change. New information, either empirical or theoretical, may cause us to revise as well as to fill in gaps in our ideal theory. I have argued that having an ideal is necessary to guide our progress but that we may be prevented from agreeing on a complete picture of an ideal. Coming to incompletely theorized agreements lets us start with an incompletely theorized picture of our ideal and then revise it later.

Sunstein's method of coming to incompletely theorized agreements will be an important component of constructing our ideal theory, as I explain below. But his kind of incompletely theorized agreement is not entirely suitable for our task. This is because Sunstein's is a theory about the law and about judicial practice; it's supposed to explain how judges who have different theories about the law can still come to agreements in particular cases. This explains Sunstein's reluctance about conceptual ascent (1760-1). Critics of Sunstein's view can argue that conceptual ascent is an important part of theorizing—that when we start with incompletely theorized agreements, we will frequently need to agree on some theory in order to make decisions. As this objection has it, "seemingly similar cases provoke different

reactions, and it is necessary to raise the level of theoretical ambition to explain whether those different reactions are justified…" (Sunstein 1760). In a tricky case of killing which is neither clearly unjustified murder nor clearly justified self-defense, we may need to appeal to one theory or another of the wrongness of murder in order to determine how to treat the defendant. According to these critics, then, "A distinguished judge will seek to add a good deal in the way of both width and depth by exploring other cases and by deepening the theoretical ambition of his analysis. He will therefore experience a kind of conceptual ascent in which the more or less isolated and small low-level principle is finally made a part of a more general theory" (Sunstein 1760-1). We must start with incompletely theorized agreements and then build off of them.

Sunstein responds that these critics have ignored "some of the distinctive characteristics of the arena in which judges must do their work," such as the complex legal and judicial system of which judges are only a small part (1761). In particular, the fact that judges must rely on precedent in making rulings will hamper their ability to decide according to a theory, since their decisions must reflect prior agreements made by others (Sunstein 1761-2). But these characteristics do not apply to moral theorizing. We don't have to respect precedent in moral theory, and we are not part of a legal and judicial system. Although we start with incompletely theorized agreements, we ought to try to develop as complete an ideal theory as we can from this start. Conceptual ascent will be a valuable tool for us, because we can start with particulars and build up from there.

Agreement on particulars

Earlier in this chapter, I defended the necessity of ideal theory against the criticism that we don't need ideal theory to tell us that we should make obvious improvements, such as getting rid of Jim Crow laws. I argued that while ideal theory may not be necessary for telling us how to make individual changes, particularly obvious ones, it is necessary for guiding us in making sustained progress over time. I made an analogy to climbing a mountain: you may be able to take the obvious path most of the way, but detailed knowledge of the mountain becomes necessary once you're close to the summit and need to know how to climb to the top. Similarly, we may not need to keep the ideal in mind in order to make every single improvement, but after we have corrected obvious injustices, we'll need the ideal in order to make sure we're making changes that will lead to sustained progress. When it's unclear whether or not we should make a given change to our society, or how exactly we should change it, the ideal is necessary for directing our choice between possibilities.

But how do we figure out the route to the top? In the face of the Sen constraint, we know that we may not be able to draw a complete map, both because our theories aren't filled out and because there's disagreement across theories. But the Rawls constraint tells us that we still need a map. We must work together to draw this map by coming to (incompletely theorized) agreements on as many points as we can. We draw a map that's as complete as we can make it, and then we proceed down that path. As we get closer to the ideal, we refine our agreements, and we attempt to develop theory that will allow us to extend our judgments (as I noted above, this is a departure

from Sunstein's method of incompletely theorized agreements in law). But we acknowledge that our picture of the ideal is incomplete, and so we expect that it will need revision as we progress toward it.

We begin by agreeing on particulars. As I said before, ideal theory is about identifying an ideal—it's about identifying what the best society, morality, or whatever looks like. We can agree, for example, on particular ills that the ideal society would avoid—nepotism in hiring, obvious racial preferences, hate speech, a permanent underclass, and so on.[25] As Sunstein points out, we can agree on many of these particulars without agreeing on the theory behind them (1736). We might, for example, disagree on the reason nepotism is wrong—whether it's unfair to our relatives or to the other candidates. But we don't have to agree on the cause of the wrongness of nepotism to agree that it's wrong. And the particulars don't have to be completely particular. We may agree to save endangered animals in general, not a particular owl or even species of owl. In this stage, we strive for as broad an agreement on particulars as possible. We cannot expect complete agreement on a complete set of particular facts about the ideal society—the Sen constraint tells us that—but we strive to get as far as we can.

Once we have agreed on particulars, we try to coalesce them into an ideal—we try to construct a society that has those particulars. We pay attention to the interactions between those particulars—is it possible to construct a coherent ideal that has racial

[25] As political liberals have noted, we will want to limit the participants in our incompletely theorized agreements to those who hold reasonable views, so that we eliminate straight out the possibility of agreeing on clearly immoral particulars (such as state-mandated racial segregation). As political liberals and their critics have noted, figuring out who, or which views, counts as "reasonable" is fraught. See Rawls, *Political Liberalism*; Quong*;* Gaus*;* and Enoch.

equality, freedom of speech, and no hate speech? If it's not, which of these particulars is less important? At this stage, we strive to create as complete a picture of the ideal society as we can. Here, we begin to factor in social science. What can current sociological, economic, psychological, etc. research tell us about the interactions between our different desired particulars? We will also want to consider certainty, both moral and empirical: how certain we are that the features of the ideal we've identified don't violate any moral constraints and how certain we are that they're feasible. More-certain moral judgments should have more-inviolable places in our ideal, other things being equal.

The goal is to get an ideal that's as complete as possible. In commenting on Rawls, Simmons makes a distinction between integrated and piecemeal ideals. The danger with piecemeal ideals is that "a particular policy might, for instance, be a good bet for remedying a particular injustice (or kind of injustice), while at the same time being a policy that retarded, stalled, or set back efforts to achieve overall justice" (Simmons 21). That is, it's dangerous to pursue a narrow ideal, because your efforts could worsen the chances of achieving other ideals. Recall the discussion of the problem of second best from earlier. Pursuing racial justice without thinking about gender or socioeconomic justice could make things worse in those arenas. Thanks to the Sen constraint, we know that it may be impossible to form an ideal that is completely integrated, because our ideal may always be incomplete. But the dangers of coming up with a piecemeal ideal should push us to make our ideal as integrated as we can. Recall that Sunstein was wary of conceptual ascent because of unique features

of our legal system. Not only are those features not present in our moral theorizing, but we also have positive reason to want to engage in conceptual ascent: without it, incompletely theorized agreements in one area might negatively affect the progress of justice in another.

If our ideal is incomplete, will that make it unstable? In thinking about political liberalism, Jonathan Quong worries that if the parties to the overlapping consensus don't reach agreement on fundamental principles, there will be "insufficient common normative ground" for them to use to reason with each other about the justice or injustice of policies (173). Similarly, maybe this method of starting with particulars will lead to a shaky consensus, because agreement on one particular doesn't necessarily transfer to agreement on another particular. But consensus can be stable if it is not just shallow but also broad: if it contains not just one but many points that people converge on, even if they are converging for different reasons. We can get convergence on particulars such as saving endangered animals, not having nepotism in hiring decisions, and racial equality even if the theories behind those particulars don't agree or are incomplete. We don't just pay attention to one of those particulars in constructing our ideal theory; we pay attention to all of them. The "common normative ground" Quong's worried about comes from this agreement on particulars. We will have to compromise on some features of the initial specification of the ideal, although we can try for greater convergence as we fill in the theory behind the ideal. But because we don't have to agree on theory at the initial stage, we can create broad

overlapping stability. This will give us the stability Quong thinks we might lack if we focus on a single point where we happen to converge.

What about the cases where it's impossible to come to an incompletely theorized agreement? Consider the two Supreme Court cases *Plessy v. Ferguson* (1896) and *Brown v. Board of Education* (1954). The Court decided in *Plessy* that the "separate but equal" doctrine of racial segregation was constitutional; in *Brown*, it reversed that decision. The only thing that can justify *Brown* as opposed to *Plessy* is a theory, not an incompletely theorized agreement—we need to know the theory behind why "separate but equal" is morally reprehensible. And this is only one of many examples of choices we make that must be informed by picking a theory; an incompletely theorized agreement won't be good enough.

Cases like these show us that coming to incompletely theorized agreements can only take us so far. Some choices we make when we construct the ideal seem to require us to pick one theory over another. As in the case of the competing theories of *Plessy* and *Brown,* some of these choices will be extremely important. The ways we treat different races are closely bound up both with practical matters about how we organize our society and with philosophical questions about how we view our fellow human beings. In this sort of case, we can't just put off making a decision on how to treat the races: we have to write laws for our society now.

There are several possibilities for bridging these gaps in our ideal theory. First, rules and analogies help (Sunstein 1743). If we can think of cases in which we agree unequal treatment is inherently unfair, and we can analogize those cases to the case of

race relations, we may be able to come to an agreement on race relations, even without doing conceptual ascent. Second, we may be able to use empirical work to come to an agreement that doesn't have a particular theoretical basis; we might look at empirical work that shows that separate but equal, whatever its underpinnings in theory, is in fact unworkable in practice, because separate facilities are never in fact equal. Finally, since our goal is to come up with an ideal that is as complete as possible, we could check the competing possibilities for coherence with the ideal we already have. Which model of race relations, *Plessy*'s or *Brown*'s, better accords with the rest of the incompletely theorized agreements comprising our ideal? If we come to obvious, easy incompletely theorized agreements first, we may find that only race integration coheres with the rest of our ideal theory.

If none of these work—if we can't analogize this situation to others, empirical facts give no guidance, and either option would cohere with the rest of our ideal—then our ideal will have to remain silent on this question, for now. Perhaps later, when we've come to more incompletely theorized agreements and have developed more theory, we can revisit this question and get an answer. But for the time being, we'll have to operate using an ideal that has a gap in it. This is a defect in this kind of ideal theory, particularly when the gaps are on central issues such as race relations. But given the Rawls and Sen constraints, that we need an ideal to guide our progress and that the ideal is going to be incomplete, this is the best we can do: as Simmons says, we will have to "muddle through the best we can" in this kind of case (24). Some

areas of law and policy will likely have to be made without an ideal in mind; we can integrate them into the ideal later as the theory develops more fully.

Transitional and non-ideal theory

Once we have a draft of our ideal, we must do transitional theory: we must figure out whether we can get there, and if so, how. Empirical work, such as the work done in the social sciences, will help us to figure out the possible paths we can take toward the ideal. The use of economics, psychology, political science, and so on is clearly necessary in the construction of transitional theory, as Simmons (for example) notes (19). Is the ideal feasible, or does it seem like it will be feasible in the future? What pathways can we take there? Will they be straightforward or more circuitous? Is the ideal feasible with the resources we have now, or only with additional resources' development in the future, or even if we suffer resource losses between now and then? When we achieve the ideal, is that achievement likely to be stable, or will it quickly collapse, making it better for us not to try to achieve it? Social scientists know that their research is open to revision and debate, just like research in the other sciences, but they can still help us with answers to these questions. In addition, there is a moral aspect of transitional theory. All other things' being equal, should we improve economic outcomes for people before we improve their social standing, for example? Social science may be able to help us here too, but ultimately some of these will be normative questions we will need moral and political theory to answer.

While we're doing transitional theory, we must engage in a related task: we must do non-ideal theory. We will have obligations with respect to the transition to the

ideal—to save more money, perhaps. But we also have obligations that have nothing to do with the transition. One example, which I discuss at greater length in another chapter, is poverty. We have obligations to relieve the suffering of the very poor whether or not doing so aids our transition to the ideal world. Civil disobedience is another example. We may have obligations to resist unjust laws whether or not doing so speeds up the removal of those laws. The development of this kind of non-transitional non-ideal theory is a distinct task from the development of transitional theory, but its results will bear on our transitional theory. If the transitional theory we develop conflicts with our non-transitional non-ideal obligations, we will have to rethink at least one of these theories. A transitional theory that requires us to save all our money in order to spend it on infrastructure at a later date would conflict with our obligation to assist those in severe poverty, and so it would be impermissible to adopt this theory. We would have to find a new way to get to the ideal.

Revising the ideal

Once we have our path to the ideal set, we embark on the transition to the ideal. But—and this will be obvious from the procedure I've described—our ideal is still incomplete. We have nowhere come up with a complete theory behind the ideal; in addition, the set of particulars that compose our ideal may not be complete. This means that revision and adjustment of our ideal and of the pathways to it will almost certainly be necessary.

As we proceed along the path toward the ideal, we revise it in two different ways. First, our ideal and the theory behind it become more complete. Advances in our

empirical knowledge will help us to fill in gaps in our ideal—we can see whether the particulars we've come up with can coexist. We will also fill in gaps in our moral knowledge. Slavery used to be a contentious moral issue; now it's not. We can expect moral progress on other contentious issues as well. We saw above that Sunstein is wary of completely theorized judgments, because they may be rigid in a way that prevents us from changing our minds as we evolve morally (1749). If we fill in our ideal theory incorrectly, we may encounter this problem. That's why we start with incompletely theorized particulars. But we gain in coherence and consistency when we fill in the theory, as Sunstein also notes, and so we should strive to add theory to our incompletely theorized agreements when we can (1761).

Second, and relatedly, we will surely realize that some of our initial specifications of the ideal were wrong. We might have been blinded by biases or simply not had all the facts about what really would be best. Or, as Sen puts it, we may discover that initially plausible general principles may become implausible when we discover that they conflict with other initially plausible general principles (107). This means that our ideal will not just become more filled in; it will also change. And as it changes, our path to it will also change. We may find ourselves having to backtrack, just like when I take the wrong path when climbing a mountain. But as the Rawls constraint tells us, we need the ideal in order to guide our progress. Without an ideal—if we just hop from a worse society to one that seems better—we are even likelier to realize we need to backtrack even more significantly.

The process I've outlined is sketchy—it's a heuristic, not a guarantee. But within the constraints imposed by Rawls and Sen, it's the best we can do. If we were working in ideal conditions—if we lacked the burdens of judgment—we could possibly expect an ideal theory that would be both theoretical and practical—it would be complete and true *and* could guide our progress right now. But because we're in non-ideal conditions, this process is one way to develop ideal theory that can guide our progress to a more just society right now. This doesn't mean that we should stop engaging in theoretical ideal theory, and we should let it inform the ideals we use to guide our progress over time—but we should keep in mind that our efforts here so far have been incomplete, and that that incompleteness currently looks to be a lasting feature of ideal theorizing. We can hope for better, however. The process of practical ideal theorizing requires constant revision of the ideal and the path we take to get there. As we continue to refine the ideal as we're making progress, we can hope to find that we approach the theoretical ideal after all.

Chapter 1, in part, is currently being prepared for submission for publication (as "Incomplete Ideal Theory"). The dissertation author was the primary investigator and author of this paper.

Chapter Two

Ideal Theory and "Ought Implies Can"

There is an apparent tension between two tasks of moral theory. On one hand, moral theory is supposed to provide a standard for us to live up to. Call this *unyielding* moral theory, since it does not yield to facts about our individual psychological or motivational shortcomings. But at the same time, morality has to provide action guidance for us as we are, with our flaws and idiosyncrasies. Call this *yielding* morality: in order to guide our actions, morality must yield to what we are like. These tasks are in tension when we cannot live up to the standard morality prescribes for us. How can morality guide our actions then?

This tension is related to another debate, about the correct meaning, plausibility, and role of the *voluntarist constraint*, that *ought implies can***.** If ought implies can, as is commonly believed, we cannot be obligated to do something that is impossible for us to do. But there is disagreement about what the constraint means. How strictly should we interpret the "can" in "ought implies can"?

In this chapter, I argue that we can use the distinction between *ideal and non-ideal theory* to make progress on these two issues. My argument has two strands. First, I look at moral theories that are unyielding, yielding, and moderate. Each has its problems. Unyielding theories are often criticized for being too utopian and unrealistic. Yielding theories are liable to seem too complacent. Some moderate theories simply split the difference between these two extremes. But these moderate theories may correctly diagnose the problems without necessarily solving them. I also

develop a parallel argument about political theories: these, too, we can sort in terms of how yielding they are.

In the second strand of argument, I interpret and assess the voluntarist constraint. I argue that there are different modalities appropriate to different sets of obligations and, hence, different readings of the voluntarist constraint. That is, there is no single "can," or "ought," in "ought implies can." We should recognize the ways in which different kinds of inability constrain an agent's obligations.

To bring these two strands together, I argue that different moral and political theories employ different versions of the voluntarist constraint: unyielding theories use a thinner version of the constraint, while yielding theories use thicker ones. We can resolve the tension between unyielding and yielding moral and political theory by distinguishing different interpretations of the voluntarist constraint. To do this, we should appeal to the distinction between ideal and non-ideal theory. Ideal moral or political theory identifies the best standard, without yielding to our flaws, and so its demands are constrained only by a thin conception of physical capacity. In contrast, non-ideal moral or political theory tells us what to do given our present situation and shortcomings. Its demands are constrained by thicker conceptions of our capacities and opportunities which yield to features of our psychology. As we make progress from non-ideal to ideal, we thin out the voluntarist constraint.

I close by considering what all of this tells us about ideal and non-ideal theory. I consider the relationship of this ideal/non-ideal distinction to the one Rawls lays out. I argue that the theory I have developed gives us good reasons to see feasibility as

wholly scalar rather than binary. Finally, I show that ideal theory and non-ideal theory can, in cases such as the one I describe in this chapter, exist on a continuum. Here is one instance where considering how the ideal/non-ideal distinction can be applied will tell us something about the theoretical aspects of the distinction.

2.1 Moral theory

Should our moral theories be aspirational, or should they instead be constrained by our concerns, projects and loyalties? I have called this the tension between unyielding and yielding moral theory. Both of these options have compelling intuitions behind them. It makes sense that morality wouldn't yield to us. What's right is right, our flaws notwithstanding. On the other hand, it makes sense that morality should set a standard appropriate for our motivations and limitations. Morality should tell us how to guide *our* actions, how to live *our* lives. In this section, I consider these different ways of understanding moral theory and the difficulties that they face.

My goal is to show that all of these moral theories have their problems. The moderate moral theorists are right: unyielding moral theory as it exists is too demanding; yielding moral theory as it exists is too lax. But while moderate moral theory correctly *diagnoses* the problems with its competitors, I will argue that it does not correctly *solve* them. The moderate moral theorist is unable to accommodate both of the compelling sets of intuitions underlying the extremes. Instead, moderate moral theory will wind up getting pushed to one or the other extreme.

Unyielding moral theories

Think about the extraordinary demands act utilitarianism places upon us. The classic example here is Peter Singer's view about beneficence. Because I must maximize utility, I must give money to poverty-relief organizations right up to "the point of marginal utility" (Singer, "Famine" 234). That is, I must give until further giving would make me worse off than those I am aiding. As Singer notes, this would fundamentally change my life, and the lives of all others not living in dire poverty ("Famine" 234). No more vacations or coffee shops or hobbies. No Christmas presents for my children. No choice in a career—I have to make as much money as I can in order to have more money to give.[26] There are exceptions to all these rules, *if* taking a vacation or giving Christmas presents or picking a career I like turns out to be the best way for me to reduce suffering. But in almost every case, I will best reduce suffering by giving away almost all my money. And this donation has to be to the most effective charities around, so no donations to the ASPCA or EMILY's List or the Republican Party, whatever organizations I feel best reflect my beliefs and priorities.

It's easy to see how act utilitarianism comes in for criticism as an excessively unyielding moral theory. There's no room in its demands for *me,* for the projects I care about and the causes that are important to me. This complaint expresses a sense that act utilitarianism runs afoul of our psychology. Central to what we are like is our concern for ourselves and those who are close to us. It seems important to a lot of us to have ongoing projects that we pursue even when they don't maximize utility.

---

[26] Singer expands on this in his recent book *The Most Good You Can Do.* Websites such as 80,000 Hours, which, like Singer, is affiliated with the effective altruism movement, help individuals to figure out which career will allow them to maximize utility. Being a hedge-fund manager might be better, if you are capable of donating a very large percentage of your income, than being a teacher or an artist or even an aid worker.

Choosing a career is for most of us a decision about what we're good at, what's important to us, and what we like to do, not about what will allow us to rake in the most cash to give away. Act utilitarianism seems to ride roughshod over all of this.

But of course act utilitarianism is not the only unyielding moral theory. Deontology may let me favor loved ones over the distant needy. But because deontology issues absolute prohibitions on certain action types, I might have to sacrifice a central life project if carrying it out meant telling a lie. Or think about the psychological difficulties associated with figuring out and acting on the virtuous mean between two vices. Aristotelian virtue ethics is unyielding too.[27]

Of course, none of these theories necessarily makes extreme demands. Many deontologists reject the idea that it is always wrong to lie; some consequentialists have worked to find ways in which consequentialism can be less demanding; some virtue ethicists are sensitive to the criticism that virtue ethics holds us to the too-high standard of behaving as perfectly virtuous beings would.[28] And yet consequentialism, deontology, and virtue ethics are unyielding. They all set standards that do not take our individual flaws into account.

Yielding moral theories

On the other end of the scale are moral theories that yield to what we are like. These theories tie their commands to our individual natures and limitations, aiming for a set of moral demands that can be integrated into our lives as they are. In much of his

[27] The situationist critique of virtue ethics suggests that the virtues are even rarer than we might have thought. This gives us even more reason to think that a moral theory that requires us to possess these virtues fails to yield to facts about us. For a prominent version of situationism, see Doris.
[28] On this critique of Kant, see Korsgaard; on consequentialism, see Brink (256) and Railton (148-56); on virtue ethics, see Swanson.

work, Bernard Williams defends a view of ethics on which our duties, whatever they are exactly, are keyed to our individual motivational abilities.[29] Our ground projects, whether ballet or our family life, are what give us reasons to keep living (Williams, *Moral Luck* 12).

Unyielding morality, on the other hand, alienates a person from his own projects because it forces him to give up the projects that are the expression of his convictions whenever they conflict with the rules of morality (Williams, "Critique" 116-17). Because more overall utility is brought about by reducing global poverty or by working to stop climate change than by having a ground project of composing beautiful sonatas, the utilitarian must give up his ground projects. If composing a beautiful sonata means breaking a promise, the deontologist must give up her ground projects.

Similarly, Wolf is glad not to know any moral saints, who spend so much time being virtuous that they have no time for hobbies, culture, or having a sense of humor (421-22). The people we actually like being around pay some attention to morality, but they also have some interests or traits "that have low moral tone. In other words, there seems to be a limit to how much morality we can stand" (Wolf 423). This suggests that morality is neither the only thing nor the most important thing in our lives. "Our values cannot be fully comprehended on the model of a hierarchical system with

[29] Some of Williams's skepticism about unyielding morality may spring from his internalism about reasons, but it doesn't have to. We might think that there are external reasons to comply with morality and yet maintain that morality may not ask us to do things that would make our lives not worth living. As Alex Worsnip noted in correspondence, Williams's (and, later, Susan Wolf's) views are mainly negative, *against* unyielding views of morality—it is interesting to note that they themselves do not offer positive yielding moral theories.

morality at the top"; rather, we must consider both what it would be moral to do *and* what kind of lives would be good lives for us to live (Wolf 436-8). Morality will at least sometimes give way to other things that matter to us.

Wolf and Williams are guided by the same basic intuition: strictly observing the rules of unyielding morality, whatever they turn out to be, will be inimical to the possibility of having a life worth living. Williams says that we should instead start with ourselves. What are my projects? What would a good life for me look like? A conception of morality that is to have authority over us must respect our lives and projects. We have moral obligations, but they must give way when they conflict with the projects and interests that are important to us. Yielding moral theory provides us with a space to pursue what we care about most, a space morality cannot take away from us.[30] The theory is yielding because it insists that morality make room for with the ground projects that are central to our psychology—the source of our loyalties and concerns.

Yielding moral theory still makes demands on us. The projects that give shape to our lives may themselves be difficult, or it may be hard for us to figure out what our projects really are. But once we know what those projects are, impartial morality is not allowed to intrude on them. Because these concerns constrain the reach of morality, a yielding moral theory is less demanding and more easily integrated into our lives.

Moderate moral theories

[30] Although Williams does note that some of us may have a ground project that consists in complying with impartial morality, in which case morality and our ground projects do not conflict (*Moral Luck* 12). Similarly, Wolf doesn't think it's impermissible to pursue moral sainthood (435).

Unyielding moral theory is intuitively plausible if we think that a moral theory can and should hold us to high standards. If we fail to live up to its aspirations, so much the worse for us. On the other hand, yielding moral theory is intuitively plausible because it seems that morality shouldn't require us to do things that are beyond our reach. If morality's demands are unrealistic, so much the worse for that conception of morality. Moderate moral theory attempts to offer an alternative that bridges this gap.

Consider Owen Flanagan's objections to Williams. Flanagan claims that as we develop psychologically, we come to be able to "override [our] natural partiality, at least up to a point," and moreover, we can come to see good reasons for doing so (71). Williams may be right to think that our projects are important to us, but he's wrong to give them an all-consuming significance. While Flanagan concedes to yielding morality that we must acknowledge that people have their own points of view, projects, and partiality to these projects, he refuses to concede that this gets us any "categorical limit" on morality (101).

But Flanagan is also critical of unyielding moral theories. Virtue ethics poses a problem because we can't come up with a complete list of virtues that don't conflict with each other (Flanagan 33). Deontology is problematic when it requires us to abstract from our particular relationships with others (Flanagan 88). And act utilitarianism is problematic because it is impossible for us to, with our limited time and cognitive resources, continually compute all possible actions we could take, their consequences, and the relative utility of each of these consequences (Flanagan 33-4).

All of these unyielding theories require us to do something that is impossible given our psychological makeup.

Thus Flanagan develops a Principle of Minimal Psychological Realism: "Make sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behavior prescribed are possible, or are perceived to be possible, for creatures like us," that is, possible given our psychology (32). This is meant to be a constraint on moral theorizing. But it's a constraint that neither the unyielding nor yielding moral theorist has good reason to accept. An act utilitarian can hold her ground. If our psychology makes us prone to immorality, it is our psychology, not morality, that should give way. After all, no one ever guaranteed us an easy or comfortable morality. And a yielding morality might insist that minimally psychologically realistic theories cannot guide individuals' actions when, after all, each individual has her own reasons and motivations. If my motivations conflict with what is minimally psychologically realistic for me, how can I act morally and still have a life worth living? When moderate moral theory rests its conclusions only on what our psychological abilities are, it cannot convince opponents who have a different view of what abilities our morality should be keyed to. A moral theory that splits the difference between yielding and unyielding will not work. So far, we have not seen a good way to resolve the tension between unyielding and yielding morality. Resolving the tension will require us to consider the ambiguity of the voluntarist constraint. First, however, we should consider a similar problem in political theory.

2.2 Political theory

There is a political debate that runs parallel to the one in moral philosophy. While Flanagan's moderate position comes as a reaction to defects he perceives in unyielding and yielding moral theory, the debate develops in the opposite direction in political philosophy. Rawls presents a moderate version of ideal theory in *A Theory of Justice;* subsequent work has criticized him both for being too yielding (such as Cohen), and for being too unyielding (such as Anderson, Farrelly, and Mills). In this section, I begin by explaining what's moderate about Rawls, and then I explore these criticisms (that he is excessively yielding first, that he is excessively unyielding second).

Moderate (ideal) theory of justice

I have discussed Rawls's ideal/non-ideal distinction in much more detail in Chapter 1, but there are several features of Rawlsian ideal theory that are especially relevant here: the limitations present in the ideal, the circumstances of justice, and the focus on human beings. In Chapter 1, I considered a puzzling feature of Rawls's account of the ideal/non-ideal distinction—that it's not clear whether natural limitations and historical and social contingencies are part of ideal or of non-ideal theory. Rawls points out that we should expect certain restrictions on liberty even in a "well-ordered society under favorable circumstances" (that is, in the ideal); these restrictions on liberty include regulations about the liberty of thought and conscience, as well as restrictions on children's freedom (*Theory* 215). However, some limitations and contingencies seem distinctly non-ideal, as in the case of a society which is too resource-poor to reach the ideal (Rawls, *Law* 5). I suggested that the solution is to see

the distinction in terms of remediability. Those limitations and contingencies that cannot be remedied *in any set of circumstances* constrain ideal theory, whereas it is the principal job of non-ideal theories to remedy those limitations and contingencies that can be fixed. The natural limitations that remain are likely not to present much of a problem for ideal theory; we will restrict the liberty of children even in the ideal world.[31] Yet they still have a moderating effect on how ideal the ideal can be. The ideal must be constrained by certain facts about human life, even though those facts are minimal and apply to any human society.

Relatedly, Rawls argues that justice is only possible under certain conditions (what he calls the *circumstances of justice*). For example, justice is only possible because people coexist at the same time in the same place; people are roughly physical and mental equals, such that no one person can dominate the rest; people are vulnerable to attack and to their plans' being blocked; people have their own life plans which lead them to make conflicting claims on resources; people sometimes lack knowledge or judgment; religious and philosophical doctrines are diverse (Rawls, *Theory* 109-110). Perhaps the most important of the circumstances of justice is the condition of moderate scarcity: "Natural and other resources are not so abundant that schemes of cooperation become superfluous, nor are conditions so harsh that fruitful ventures must inevitably break down" (Rawls, *Theory* 110).

[31] It seems like there will be only limitations, no contingencies, in the ideal, since contingencies are by definition remediable in at least some sets of circumstances.

Because the circumstances of justice are required before we can develop any principles of justice at all, that means that these circumstances are present in both ideal and non-ideal theory. But notice that that means, in effect, that ideal theory is not the most ideal theory out there. Wouldn't it be more ideal if moderate scarcity didn't exist—if we all got everything we wanted all the time, rather than having to work for it? Or wouldn't it be better if we were all able to figure out the correct religious and philosophical doctrines and agree on them? This has a moderating effect on Rawls's ideal theory. A more unyielding ideal theory might say that ideal justice is only possible in conditions of no scarcity.[32]

And finally, as the introduction of the circumstances of justice makes clear, Rawls's ideal theory is ideal theory *about humans*—about beings who have needs, can make plans, have incomplete knowledge and imperfect judgment, are capable of cooperation, and so on. Rawls does not have a particularly dim view of human nature, as we see in his apparent belief that full compliance with the principles of justice is possible (*Theory* 216). His ideal theory is not particularly yielding—it is optimistic about what humans are capable of. But it is also not as unyielding as it might be, because it limits itself to laying out an ideal *for humans.*

Unyielding (ideal) theory of justice

[32] For this to be an ideal theory *of justice,* we would have to hold that justice is possible and necessary even in conditions of no scarcity. I suspect there's an argument to be made for this position: given what we know of human nature, it seems likely that people would covet and steal even if they gained nothing by doing so. Maybe this is plausible, maybe not. But a more unyielding ideal theory *of justice* at least seems like a possible position to take, rendering Rawls's theory relatively moderate.

It might seem obvious that ideal theory should be ideal theory for humans—after all, that's whom we're making rules for, right? Surely justice should take account of, at least, some very basic facts about human nature. But not everyone takes that position. GA Cohen may be the most well-known exponent of the more unyielding view that justice is in no way constrained by what humans are like. Cohen offers a *fact-principle argument* for his view that ultimate principles of justice are fact-free (229-73).

Suppose I believe that it's wrong to eat meat. What reason could I have for holding this belief? Because of the fact that eating meat causes unnecessary pain to animals. But we can ask a further question: *why* do I believe that this fact causes meat-eating to be wrong? There must be a further principle behind this fact, perhaps the principle that it is wrong to cause unnecessary pain to animals. But why should we accept this principle? Perhaps because of the fact that animals are the kind of beings who have an interest in not being in pain. But again, we don't end with a fact. *Why* is this fact important? It must be because of the further principle that we should not cause unnecessary pain to beings who have an interest in not being in pain (or something like that).[33]

Cohen argues that this fact-principle sequence will always bottom out in a principle, not a fact. You can always ask *why* a fact is important, and this explanation of *why* will always be a principle. At some point, there will be an ultimate principle

[33] This is a metaphysical, not an epistemic, claim: the fact-principle argument is subject to the "clarity of mind requirement: the argument "applies to anyone's principles, be they correct or not, *so long as she has a clear grasp both of what her principles are and why she holds them"* (Cohen 233). It is an argument about what actually grounds principles, not about what fallacious reasoners happen to believe grounds principles.

(or principles) that cannot be explained by a further fact (perhaps this principle [or principles] is self-evident, or perhaps it is justified by a non-normative principle) (Cohen 238). But the ultimate principle will be fact-free. If you don't believe this, Cohen says, then just try to explain how some "credible and satisfying explanation of why some *F* supports some *P* invokes or implies no such more ultimate principle" (236). It doesn't seem like we can explain why facts matter without showing how they matter for principles.

If Cohen's right about the fact-principle argument, then ultimate principles of justice are fact-free. And if this is right, then we can see how Cohen's theory makes Rawls's look moderate. The ultimate principles of justice cannot be constrained by any facts *whatsoever,* including even the very most basic facts about what humans are like or what they are capable of. Cohen does allow for *rules of regulation,* which *are* sensitive to the facts—these rules take into consideration stability, publicity, and so on in addition to justice (286). But these rules of regulation are derived from the ultimate principles of justice (Cohen 275). "It is wrong to eat meat" is a rule of regulation that relies on a fact-free ultimate principle, something like "We should not cause unnecessary pain to beings who have an interest in not being in pain."

This doesn't mean that Cohen denies the voluntarist constraint. On the contrary, he accepts that what *we* ought to do is constrained by what *we* can do (Cohen 250-54). But what *we* ought to do is determined by rules of regulation, not by ultimate principles of justice. All ultimate principles, Cohen says, have a conditional form: "One ought to do *A* if it is possible to do *A*". But it's important to see that while the

application of the ultimate principles is constrained by facts, their *content* is not. As Cohen says, no facts can refute ultimate principles (251). Thus he accepts the voluntarist constraint while also denying that facts (about possibility or anything else) have any role to play in the formulation of ultimate principles.

True justice, then, is to be found in the ultimate principles, not in the rules of regulation. Given facts about human life—about certain practical problems we face—equality is an infeasible policy goal, Cohen writes: "one can only approach it, but that is not in my view a reason for identifying justice with whatever workable rule comes closest to equality, as opposed to with what we are trying to approach, that is, equality itself" (279). Perhaps only angels, or nobody at all, can achieve the true justice spelled out by the ultimate principles, whatever they are. But rather than saying "so much the worse for the ultimate principles," Cohen responds: "so much the worse for people!" True justice may be (in fact at one point Cohen claims straightforwardly *is*) infeasible for people, but that doesn't make it any less true justice (254).[34]

Yielding (non-ideal) theory of justice

The previous two sections focused on differences in defining *ideal* theory of justice: while moderate Rawls argues that this theory should be constrained by facts, unyielding Cohen thinks it should be constrained by none whatsoever. But there's a third alternative: yielding theory of justice. This alternative is couched in *non-ideal* terms, to the complete exclusion of ideal theory. Critics of ideal theory—Anderson, Farrelly, Mills, and others—charge that it is precisely ideal theory's ignorance of the

[34] See also Cohen's discussion of the circumstances of justice (331-36) for another way in which he rejects Rawls's moderate theory.

facts that makes it pointless or even pernicious. The third alternative to unyielding and moderate ideal theory of justice isn't yielding *ideal* theory of justice; rather, it's yielding *non-ideal* theory of justice.

For example, Anderson writes, "we need to tailor our principles to the motivational and cognitive capacities of human beings" (3). She writes that *just* institutions must compensate for our moral and motivational deficiencies and take account of our interests and motivations as they are (Anderson 4). Notice that she's claiming that the institutions of prescribed by ideal theory, which do not take these limitations into account, cannot be just, since they don't suit actual people. Anderson's underlying claim is that what's just is determined by taking into consideration people *as they are*, not people as they might be.

There is a similar emphasis on facts in Farrelly's critique of ideal theory. In fact, Farrelly *defines* the ideal/non-ideal distinction in terms of fact-sensitivity. Ideal theories are sensitive to few or no facts; non-ideal theories are sensitive to more. So Farrelly ranks theories of justice on a spectrum from non-ideal to ideal, where Cohen represents the extreme ideal position and Rawls is somewhere in the middle (847). Farrelly goes on to argue for theory that is somewhere on the more non-ideal end of the spectrum: in particular, he argues that we need to know the facts about what it costs to protect our rights before we can know whether we should trade some of our basic liberties for some equality of opportunity (851-52).

And finally, Mills argues that ideal theory is not just unhelpful but actually pernicious, because it ignores a particular kind of fact: the structure of power. Ideal

theory presupposes that society is not shaped by unjust conditions of social domination, coercion, and oppression, but that means that the people living in ideal conditions also do not have lives shaped by those conditions (Mills 168). Rather than coming up with theories of justice that apply to people who have the capacities and personalities that people actually do, we come up with theories of justice to fit whatever capacities and personalities people would have if they didn't live in conditions of domination, coercion, and oppression (Mills 169).

Thus a major critique of ideal theory is its insensitivity to the facts. While these non-ideal theorists do not usually attempt to construct comprehensive non-ideal theories of justice, the positive proposals they do make about justice (such as Anderson's argument for racial integration) are significantly more attuned to the facts than are unyielding and moderate theories of justice.

So we get a parallel tension in political theory to the one we saw in moral theory. In both, we are divided between those who think that a theory should make demands at (or even beyond) the limits of what people are able to do, and those who think that a theory must hew much more closely to what people are actually like, and those who are somewhere in the middle. While the philosophers arguing for each of these positions arrive at them for different reasons, I will argue that one fundamental difference between the positions is that they are each operating with a different version of the voluntarist constraint.

2.3 The voluntarist constraint

Most philosophers (although not all) accept the voluntarist constraint, that *ought implies can.*[35] They believe, that is, you can only be obligated to do things that are possible for you to do. The voluntarist constraint is widely accepted because of its intuitive plausibility. One reason for this is that the constraint seems *fair*. There's something unfair about a morality that commands that people do the impossible, and morality cannot guide our actions if it commands impossible things. But for every argument for the voluntarist constraint, opponents have developed a counterargument.[36] Even so, the constraint remains deeply compelling and seems to me likely to be true in some form. But proving this is a larger task than I have room for here. Instead, I will focus on a different question: *if* ought does imply can, *then* what is there to say about the constraint? In this section, I argue that all of the unitary voluntarist constraints philosophers have come up with are insufficient. There is no one single "can" in "ought implies can." This sets the stage for the resolution, in the next section, of the tension between unyielding and yielding moral and political theory.

Possibility

When we say that someone "can" do something, we may mean any of several different things.

Logical

---

[35] See Graham and Howard-Snyder on arguments for the voluntarist constraint. For arguments against the constraint, see, among others, Mizrahi, Vranas, and King. Sinnott-Armstrong ("'Ought' Conversationally Implies 'Can'") and Stern ("Does 'Ought' Imply 'Can'?") accept weaker readings of the constraint but argue that the traditional formulation of the constraint is too strong (see also Streumer's response to Sinnott-Armstrong).

[36] Van Someren Greve and Sinnott-Armstrong ("'Ought' Conversationally Implies 'Can'") have provided replies to the fairness argument, for example.

Some event *p* is logically possible if and only if that event is consistent with the laws of logic. Self-contradictory events are logically impossible. It is logically impossible for something to be both *p* and not-*p;* it is logically impossible for Earth to be simultaneously square and not square. It is logically possible, however, for Earth to be square. No rule of logic constrains the shape of Earth.

Metaphysical

Some event *p* is metaphysically possible if and only if there is a possible world in which *p* happens. It is metaphysically possible for the force of the Earth's gravity to be double what it actually is. There is a possible world, that is, in which gravity is twice as strong. It is metaphysically impossible for the force of the Earth's gravity to simultaneously act in different ways on two otherwise-identical objects. There is no possible world in which a physical force treats otherwise-identical objects in different ways.

Nomological

Some event *p* is nomologically possible if and only if *p* could occur given the laws of nature as they exist in our world. It is nomologically possible for me to run a mile in four minutes; nothing in the laws of nature prevents humans from being unable to run that quickly. It is nomologically impossible for me to run a mile faster than the speed of light; the laws of nature prevent anything from going faster than the speed of light.

These first three kinds of possibility were possibilities about the world: they addressed questions about what is possible given certain features of the world around

us (or of possible worlds). I will refer back to them again at the end of the chapter. The last three kinds of possibility are more immediately relevant because they are possibilities that are directly related to the capacities of human beings.

Physical

I "physically can" do something if it is possible for me to do it do it given facts about my physiology. Here, I mean solely facts about the body, as distinct from facts about the mind or brain. It is physically possible for me to touch my left elbow with my right hand. I'm strong enough and have the muscle control to bring my right hand across my body to my left elbow. It is physically impossible for me to touch my *right* elbow with my right hand; my forearms aren't that flexible.

Psychological

I "psychologically can" do something if it is possible for me to do it given facts about my psychology—that is, about what my mind or brain is capable of. If I am a psychopath with a normal, healthy physiology, it is *physically* possible for me to act as though I love my child, because I can hug her, give her Christmas presents, and tell her I love her. Someone else with my physiology (my muscles and joints) would be doing these things out of love. But for me, given my brand of psychopathy, it is *psychologically* impossible for me to love my child.

An individual's psychology places further constraints on what's possible, given that individual's physiology. A person with the psychopath's exact structure of muscles, bones, and so on, but with a non-psychopathic brain, could love her child. Our minds place additional constraints on what our bodies can do. So even though the

brain is part of the body, psychological possibility is a constrained subset of physical possibility.

Motivational

Within psychological possibility, we might make finer gradations. Consider an example, borrowed from Estlund, of someone who has a great fear of heights (230). Under normal circumstances, this person is unable to walk on a glass bridge stretching out over the Grand Canyon. It's just impossible (due to facts about her psychology, not her physiology). But if it's absolutely necessary—if it will save her life—she can muster up hidden psychological resources and overcome her fear. Once the danger is past, though, she might have the same or even a greater fear of heights. We might fairly say that what is psychologically impossible for her in ordinary circumstances becomes psychologically possible in extraordinary ones. We could say that it is *minimally* psychologically possible for her to walk on the glass bridge.

Perhaps this newfound ability has to do with her motivational structure. Within the set of things that are psychologically possible for me, it is only possible for me to be able to *want* to bring about some of those things. It is *psychologically* possible for me to keep my cool in tense situations. But it's *motivationally* impossible for me not to yell at a driver who cuts me off in rush-hour traffic: I can't keep my cool because I can't *want* to keep my cool. I mean "want" in a broad sense here—so while I do not *enjoy* taking out the trash, the fact that I do it shows that I want, broadly, to do it. We might, following Schueler, call this a "pro attitude" toward taking out the trash, even if taking out the trash is not exactly something I have a craving to do (35).

Of all the subsets of psychological possibility, I emphasize motivational possibility in particular for a couple of reasons. This divide between psychology and motivation is commonsensical. Say a bully is beating up a smaller child outside. You can't leave the house in time to save the child because of your severe obsessive-compulsive disorder: you have to perform certain rituals before you can leave the house, even though you desperately want to help the smaller child. I can't leave the house because I am extremely selfish and can't make myself want to get up out of my easy chair. Even though our leaving the house and rescuing the child is impossible in both cases, my motivational abilities seem more changeable, or up to me, or blameworthy than your psychological ones are. We might blame you for not overcoming your psychological impossibility—but I am still more blameworthy for not overcoming my motivational one. Philosophers have picked up on this commonsensical distinction. What Flanagan disagrees with Williams about is precisely whether our psychological or motivational abilities should determine our obligations.

But it may be difficult to say precisely where motivational possibility ends and other psychological possibility begins. Rather than discrete kinds of possibility, it may be more accurate to say that these represent regions on a continuum. We might go from motivations I have easily, to those which I need to strain to discover, to my everyday psychology, to things I have to strain to be able to do. And if motivational and psychological possibility come in gradations, then surely physical possibility does too. I can run a mile in 15 minutes with ease, but I would have to train hard to run it in

seven, and it is almost certainly impossible for me to run it in four. At some point, we stop talking about physical possibility and start talking about psychological possibility, and at some point we are thinking specifically about motivations and not broader features of a person's psychology—but drawing bright lines between each is probably impossible. At the end of this chapter, I will return to this issue when I talk about ideal and non-ideal theory as a continuum.

As the examples I've offered show, the different kinds of possibility will vary in how helpful they are in different contexts. If I am a psychopath, "I can love my child" will be treated as false in any context where we're considering my psychology. But when we are considering what I can physically do, "I can love my child" becomes both true and useful. But there is no single account of possibility—no one true analysis of "can." Rather, there many true analyses of "can," but which one(s) are relevant will vary by context.[37]

"Can" and "ought implies can"

We might expect those who accept the voluntarist constraint to recognize the significance of the multiple meanings of "can." Not so. [38] Consider a recent disagreement about "ought implies can" within political philosophy. David Estlund interprets the "can" of "ought implies can" in this way: "A person is able to (can) do something if and only if, were she to try and not give up, she would tend to succeed"

[37] Other kinds of possibility may be relevant. Värynen mentions epistemic possibility ("certain kinds of information pertaining to the act are available")(302). The account I am developing is open to these other kinds of possibility as well. We could also look at a related view in semantics—see, e.g., Kratzer.
[38] Chuard and Southwood (614) mention in passing the existence of multiple kinds of "can," but they do not pursue this point. Kekes and Jay note, but also do not pursue, the point that limitations beyond physical inability might constrain obligations if the voluntarist constraint is true.

(212). This is essentially physical possibility—Estlund even includes staying awake for four days straight as an example of something we can do (213). (This puts him somewhere between Rawls and Cohen.) Estlund argues for his thinner voluntarist constraint by appealing to the case of Bill the polluter, who claims that he is not required to refrain from dumping his garbage by the side of the road because he is too selfish to refrain (rather than because he has some phobia or compulsion) (220). Estlund says that selfishness is intuitively not the kind of thing that can block a requirement; therefore, no motivational inabilities block requirements, even if they are inabilities all humans share (220). Because features of human nature do not block the requirements of justice, we wind up with a very thin sense of "can" for the voluntarist constraint.

David Wiens argues that the relevant sense of "can" is much thicker: to him, "ought implies can will (in good faith)" ("Motivational Limits" 8).[39] Wiens asks us to imagine Claudia, who makes repeated good-faith attempts to write a book but each time only writes a few pages before she gives up ("Motivational Limits" 10-11). Can Claudia write a book? Wiens thinks she can't; she's proven that it is impossible for her to complete the sequence of events that results in a finished book. On Wiens's definition of "can," "A person is able to (can) do something if and only if, were she to repeatedly make good-faith attempts to complete a sequence of acts that conduces to that thing, she would tend to do that thing successfully" ("Motivational Limits" 10). So whether Bill can avoid dumping his trash by the road depends on the reasons for

[39] Graham has a similar sense of "can" in mind; on his view, an arachnophobe cannot touch a spider (342).

his claimed inability. If Bill's selfishness prevents him from making a good-faith effort to refrain from dumping, then he *cannot* refrain from dumping; if he could, with the right good-faith effort, work past his selfishness, then he *can* refrain. While Estlund believes that moral theories ought to idealize away from human nature, Wiens requires us to incorporate at least some inabilities. (This puts him in the non-ideal-theory camp, along with Anderson and the others.)

But neither of these single readings of "can" tells the whole story. Wiens identifies an important distinction between good- and bad-faith motivational inabilities: there seems to be an important difference between someone who can't do something because she is too selfish or cruel to do it and someone who can't do something even when she tries in good faith to do it ("Motivational Limits" 16-17). Think about a version of Bill who, rather than being too selfish to refrain from dumping, instead has a more sympathetic condition. Say that he has an obsession with cleanliness, so that, try though he might, it is impossible for him to make it all the way to the dump before he has to get rid of his trash. Now it seems clearer that Bill can't refrain from dumping. Estlund simply leaves us on the hook for too much—if no part of human nature counts against our being able to do something, then our obligations will far exceed what we can do.

But then again, Estlund is motivated by trying to figure out what justice is: "those to whom we owe justice do not lose their claim on us just because it might turn out that we are not, perhaps even by our nature, disposed to deliver it" (230). If Wiens's Claudia has promised a publisher that she will write a book, then her inability

to write the book does not get her out of her obligation to write the book. When Claudia has promised, we *should* say that Claudia ought (in some sense) to write the book; with Wiens's "can," we cannot. Some opponents of the voluntarist constraint contend that it lets us off the hook for things we should be responsible for; Wiens's version is especially susceptible to this criticism.[40]

Thus both Estlund's thinner and Wiens's thicker specifications of "can" seem to run into difficulties: neither can fully account for the different kinds of obligations we have. What if we split the difference with a moderate voluntarist constraint, "ought implies psychologically can"? We saw an attempt to do something similar in the debate between unyielding and yielding moral theory. Just as in that case, a moderate voluntarist constraint cannot solve the problems with the extremes. In some cases where psychology interferes, such as the case of Bill the unselfish but compulsive dumper, it seems unfair to hold someone responsible for an inability to comply with morality. But in other cases, such as when Claudia promises to finish the book but is psychologically unable to comply, it seems entirely appropriate to hold that person responsible. A moderate voluntarist constraint rules out the second kind of case off the bat. A voluntarist constraint that can *never* give us obligations we are psychologically unable to fulfill seems too lax.

And even a moderate voluntarist constraint can only provide limited action guidance. Say that a dictate of morality is psychologically possible but motivationally impossible for you. What do you do now? Since thick voluntarist constraints are

[40] For a version of this contention, see Mizrahi. Stocker writes that the voluntarist constraint "would almost certainly be uninterestingly false if considered in light of *psychological* inability," arguing that the plausible sense of the constraint is a physical sense of "can" (311).

closely tied to your motivations, they can guide your actions in a wide variety of circumstances. Thin and moderate voluntarist constraints can't do that. They may provide long-term goals for you—they may direct you to change your motivational structure—but they cannot give you immediate action guidance given the motives you have right now. So, just as we saw with moderate moral theory, a moderate voluntarist constraint doesn't seem to have the advantages that either a thin or a thick constraint does, but it does inherit their problems. It's time to investigate an alternative strategy.

2.4 Multiple voluntarist constraints

Moral theory with multiple voluntarist constraints

Rather than trying to find one single meaning of the voluntarist constraint that can account for everything we want morality to do, we must develop multiple versions of the voluntarist constraint using the different modalities discussed above. Our moral theories (whether these are utilitarian, deontological, or virtue theories, or something else entirely) must be sensitive to multiple voluntarist constraints: "(this) ought implies physically can," "(that) ought implies psychologically can," and so on. This means that any moral theory will have multiple component theories, giving us multiple sets of obligations, all of which we are under at once. We start with a thin voluntarist constraint, "ought implies physically can." This provides us with our maximum obligations. But in order to get more immediate practical guidance, we look to components of our moral theory that use thicker voluntarist constraints. Each voluntarist constraint delivers to us a distinct set of obligations: the different versions of "can" give us different sets of "ought." When it is appropriate to consider the

psychological version of the voluntarist constraint, our obligations are drawn from the set of things we are psychologically able to do. When it's appropriate to use the physical voluntarist constraint, our obligations are drawn from the set of things that are physically possible for us.[41]

This is one way to apply the ideal/non-ideal distinction to moral theory. In the ideal world, we would lack the limitations we have in the non-ideal world. We wouldn't have the psychological or motivational inabilities that make it in some sense impossible for us to comply with morality's demands. But in the actual world, these inabilities make it impossible for us to comply. That's why we need distinct non-ideal theory that we can use to guide our actions in the non-ideal world. We derive our non-ideal obligations from our ideal obligations, but we modify them in the light of the thicker senses of possibility. An example may help here.

An example: act utilitarianism

Assume that act utilitarianism is true.[42] Then think about our duties of beneficence. On act utilitarianism, I would have (at least) three sets of obligations to the distant needy.[43] First, I would be obligated to aid the needy as far as is physically possible. While this level of possibility is often unhelpful for providing immediate

[41] Notice the contrast with a related distinction, between deontic and evaluative kinds of "ought." See Howard-Snyder (1) and Chuard and Southwood (601) here. While evaluative "ought" statements ("Life ought to be fair") do not make demands on us, deontic "ought" statements ("I ought to tell the truth") do. All versions of "ought" on the non-ideal-to-ideal continuum make demands of some kind or another, so the ideal/non-ideal distinction is a distinction *within* the category of deontic "ought."

[42] This framework applies to any unyielding moral theory, but I'm using act utilitarianism here because of its simple structure.

[43] Although there may sometimes be a bright line between ideal and non-ideal theory, this is a case where they're on a spectrum. The different kinds of possibility may be difficult to distinguish from each other. But to measure our progress to the ideal, and for the sake of coming up with delineated sets of obligations, I have chosen three distinct points on this spectrum.

practical guidance, it can tell us our maximum moral obligations. Even the staunchest act utilitarian might admit that it is psychologically impossible to live up to act-utilitarian principles one hundred percent of the time—but as long as those principles are *physically* possible, we are nevertheless doing something wrong when we fail to live up to them. If aiding the needy consists in writing checks to Oxfam, then I am obligated to write checks to Oxfam until I reach the outer limit of what morality could ever require (perhaps up to the level of marginal utility, as Singer argues). This obligation is not conditioned by our psychological needs or our motivations: it is conditioned only by the actions our bodies can perform.

On a non-ideal theory that yields to facts about my psychology, I would be obligated to give to the needy as far as my psychology will let me. Perhaps it is psychologically impossible for me to never favor my family over complete strangers. I may be able, as much as it hurts, to divide my time and resources exactly equally, but I may still be unable to avoid giving my child or parent an extra hug or loving glance or time in my thoughts. In this case, then, my obligation would be to aid my own family first and only later turn to aiding others. What's psychologically possible can often be useful as a medium-term goal: I may not be able to do everything that's psychologically possible for me, but these obligations are less distant than things that are only physically possible.

And finally, consider an even more non-ideal theory, which takes into account my motivation. For the purposes of immediate action guidance, this non-ideal theory tells me that I have the obligation to aid the needy as far as I am motivated to. This

level, the most constrained set of obligations, gives immediate practical guidance. Right now, surely I can only do whatever I can currently motivate myself to do. So if I'm deciding how to act in the short term, I should only consider those options that are motivationally possible for me. This might mean that this non-ideal theory tells me to aid the very poor only after I have successfully completed or made progress on projects that are important to me. If having a collection of fine art is important to me, then I should spend as much money on art as I must and only then give money to the needy.

Whatever the correct moral theory is, we start with the maximum obligations of that theory and then add thicker and thicker voluntarist constraints to get closer to what we are actually able to do. The different kinds of possibility thus act as successive filters. When we move from ideal to non-ideal theory, we filter out obligations that are impossible in a non-ideal context. When we move from physical to psychological possibility, we are filtering out everything that is physically possible but psychologically impossible. This helps us to see which of our obligations are feasible in the short term.

But this filter metaphor isn't perfect. It might give the impression that since we are at every step removing obligations, surely we can never find ourselves with more obligations than we had at a more unyielding step. But this isn't true. As we move from more unyielding to more yielding theory, we may find ourselves with more duties to make up for things we (in some sense) can't do. If I have arachnophobia, perhaps I am not psychologically obligated to save a child from being bitten by a

spider. But I may have (psychologically possible) duties to make up for my failure to save the child (which was physically possible but psychologically impossible). I might have duties to visit the child in the hospital, to pay some of his hospital bills, to undergo therapy to try to treat my phobia, and so on.

Transition to the ideal

It might seem like the most yielding theory, the one that we actually use to guide our actions, requires us to do hardly anything. If I am not motivated to donate money to the poor, or save a drowning child, or refrain from saying something mean to someone, then on non-ideal theory with a motivational voluntarist constraint, I am not obligated to. Doesn't this let us off the hook too easily?

A first response is that even this very yielding non-ideal theory is not as lax as it may seem. There is an important distinction between "can't," "won't," and "unlikely to." We are off the hook for things we *cannot* be motivated to do, but we are not off the hook for things we can be motivated to do but *won't* do or are *unlikely* to do. If I can bring myself to want to take out the trash, then I must do it, even if I would enjoy doing something else. Simple laziness does not remove an obligation, even on the most non-ideal version of our moral theory. (Think of the difficulty, which Williams acknowledges, of completing some of our personal projects, even though they fit with our motivations (*Moral Luck* 13).) In order for non-ideal theory to be unable to command something, it must truly be motivationally impossible for us.

But even though motivational obligations can sometimes be demanding, sometimes they simply aren't. Maybe I truly cannot bring myself to want to give

strangers equal consideration to my family members. In this case, the non-ideal moral theory which uses a motivational voluntarist constraint cannot require me to give equal consideration to strangers. This would be a problem if this were the only moral theory we had. But it's not. Consider two kinds of non-ideal theory.[44]

The first kind of non-ideal theory, *non-transitional theory*, tells us about our obligations in the non-ideal world. When I've been talking about non-ideal moral theory (about the parts of our moral theories that operate with psychological and motivational voluntarist constraints), I have mostly been talking about non-transitional theory. Even if we can never become motivationally capable of doing more than we can do right now, we have some obligations we can act on now.

The second kind is *transitional* theory. This kind of theory tells us about our obligations to transition to the ideal. A common view of punishment helps to illustrate this distinction. We punish for non-transitional non-ideal reasons: in the non-ideal world, people do the wrong thing, and we want to express society's disapproval. We also punish for transitional reasons: we want people to learn from their mistakes in order to do better next time.

In the case of moral theory, I have obligations to bring my motivations in line with what is psychologically and, ultimately, physically possible for me. These are real obligations for me, because there's some sense in which I can do them—they don't disappear just because of my motivational or psychological defects. This helps to answer the worry that motivational moral theory is not demanding enough. I should

[44] For more on the transition from non-ideal to ideal, see Simmons.

try to become the kind of person who is motivationally able to comply with these other obligations. I should try to rearrange my motivational structure to comply with the obligations that are psychologically possible for me; I should try to improve my psychological structure in order to comply with the obligations that are physically possible for me. If I don't, I am ignoring some of my moral obligations. Because moral theory with a motivational voluntarist constraint is not the entirety of our moral theory, we are not reliant exclusively on a yielding moral theory.

There are complicated questions about how this transition works, but many of these will depend on whichever unyielding moral theory turns out to be true. Think about the possibility of conflict between obligations at different levels of a moral theory. All-things-considered guidance isn't easy to come by when a non-transitional non-ideal obligation conflicts with a transitional one. Say that I could either give a small amount to Oxfam (consistent with my motivationally possible obligation) or I could give no money to Oxfam at all, instead spending it on a therapist who will help me overcome my selfish motivational structure. That will make it motivationally possible for me to give more money in the long run. In this case, there is a conflict between my non-transitional and transitional obligations.

This is one instance of a broader problem of transition in ideal and non-ideal theory in general. The details about when and how we transition to the ideal will often be quite complicated and depend on the moral theory we accept. If we are utilitarians, questions about transition will be solved by whatever brings about more utility in the long term. I might be required to spend my money on therapy rather than Oxfam, so

that once therapy is over I can give much more money. If we are deontologists, our transition to the ideal will be constrained by other features of morality, such as rights. I am probably not allowed, all things considered, to murder someone now, even if that somehow expands what's possible for me later on. When obligations conflict, action guidance will vary depending on several factors, including which moral theory I accept, what my options are, and how they help or impede my transition to ideal morality. There may be all-things-considered action guidance for every moral question, but we will need to know more about the specifics of moral theory, not just its structure, to know what that guidance is.[45]

What kind of ideal?

Dividing moral theory up into ideal and non-ideal versions changes how we decide which moral theory is correct. We saw that Flanagan and Williams dismiss unyielding moral theories by arguing that some of their demands are psychologically or motivationally impossible. These are no longer good grounds on which to dismiss an ideal moral theory. When we are deciding which ideal theory of morality (ideal consequentialism, deontology, virtue ethics, etc.) is correct, we cannot bring in facts about what is psychologically or motivationally impossible for people to do. Act consequentialism cannot be dismissed on the grounds that it is psychologically impossible for people to figure out the ideal action to take (Flanagan (33-34)). Kantian deontology cannot be dismissed on the grounds that it is impossible for people to act solely from a motive of duty (Flanagan 36). If people are physically capable of doing

[45] Or perhaps we can't always get all-things-considered guidance, if moral dilemmas are possible. But I won't pursue that issue here.

these things, then ideal morality can require them to do so. Questions about what our obligations are given psychological and motivational inabilities are now questions for non-ideal theory.

But this doesn't mean that the most extreme versions of moral theory are necessarily the right ones. *Unyielding* moral theories should not be identified with *extremely demanding* moral theories. Moral theory may make less extreme demands not because people are unable to comply with extreme demands but because morality *should not* make extreme demands even on people who can comply with them. For example, consider Samuel Scheffler's moral theory.

Scheffler rejects consequentialism with an appeal to personal integrity that draws on Williams's objections discussed above (*Rejection* 9). People have particular points of view, giving them their own concerns and projects. They care about these projects out of proportion to how impartially good they are—out of proportion to how well they do at maximizing consequences. But consequentialism requires us to do whatever will bring about the best consequences. This means that we must allocate our time and energy "in strict proportion" to how we can bring about the best consequences. We must give short shrift to our own projects and concerns or even abandon them entirely.

So Scheffler winds up with a moral theory that is less demanding than act consequentialism: he defends agent-centered prerogatives, which allow agents to assign their own interests, projects, and loyalties normative significance that is out of proportion to their impersonal value (*Rejection* 20). This kind of moral theory avoids

alienation because people are not forced to give up their personal projects whenever they conflict with maximizing good consequences.

This may sound like an effort to perch in between yielding and unyielding moral theory, which I rejected when Flanagan did it.[46] But Scheffler's moral theory does not need to rely on any claims about inability. Even if people are psychologically and motivationally capable of becoming act consequentialists, morality, even ideal morality, should not require them to do so. Scheffler's theory is less demanding than consequentialism not because of claims about what we *can* do but because of claims about what we *should* do: what a good life for us is like and what is properly important to us.

Other moral theories reduce their demands for the same kind of reason. It's not that lying is sometimes permissible for a deontologist because it's psychologically impossible never to lie; instead, it's that the right moral theory will permit even ideally honest agents to lie sometimes. Ideal moral theories can come with more or less extensive demands. What they have in common is that they arrive at those demands not by looking at what people are capable of but at what morality may reasonably ask of people, regardless of their capabilities. Then supererogatory action is any action that goes beyond what morality can reasonably demand of us, not necessarily action that is beyond the capabilities of most people.

Blame

[46] To be fair, Scheffler sometimes makes claims about what is psychologically possible for us (see *Human Morality* 68). But he need not rely on those claims. He can construct an ideal theory based only on claims about what morality ought to *ask* of us, not what we *can* do.

I have a friend who is notorious for lateness and flakiness. Phone calls regularly go unreturned; he often shows up to social events late and leaves early; sometimes he cancels at the last minute. He's a dear friend, but he's totally unreliable. Those of us who have known him for years recognize this as just a part of his personality—you can't depend on him to show up on time, and you have to plan for that. But that doesn't mean it isn't frustrating. When my friend shows up late for something, and I get upset, I blame him for being unreliable. But it doesn't seem to me that I'm really blaming him for this particular act of irresponsibility. I might think to myself, "Yes, it's frustrating that he was late again, but you know how *he* is…" If it were some other friend, one who isn't chronically disorganized, I might blame that friend for a particular act of laziness. But with this friend, I know unreliability is baked into his character. It is, maybe, actually *impossible* for him to motivate himself to be on time. And if this is motivationally impossible, then there can be no motivationally possible obligation for my friend to be on time.

But there can be obligations at other levels of possibility. It seems to me that in this case I blame my friend for his character—for being the *kind of person* who's unreliable, for having a psychological makeup that gives him insufficient motivational resources for being on time. I blame him not for something that is impossible for him to do (since returning a particular phone call may be impossible). I blame him instead for bringing it about that that thing is impossible. And I will continue to blame him for that—for having a certain kind of motivation—until he changes. Until his character improves, it will be impossible for him to stop flaking out. But that's a different kind

of blame than blaming him for being motivated to do a particular thing, and it has a different kind of remedy. My friend must try to make the currently impossible into the possible.

Other moral theories may have trouble accommodating this kind of assignment of blame. On a yielding moral theory, we have no reason to blame my friend—his actions were perfectly in line with his motivations. On an unyielding moral theory, we must blame my friend for each individual act of lateness—he did not live up to the rules of morality. But if morality has an ideal and non-ideal structure, then we can blame and not blame in different ways. We can, for example, resist blaming someone for an individual wrong while saying that he should make it possible for himself to stop doing that wrong, maybe by taking a time management class. We now have a more complex notion of blame for complex situations of possibility.

Political theory with multiple voluntarist constraints

My main focus in this chapter has been on developing moral theory with multiple voluntarist constraints. But as I noted above, political theory also has a version of the tension between unyielding and yielding theory. In formulating fundamental principles of justice, Cohen is not constrained at all by what is possible for humans. Conversely, Anderson, Farrelly, and Mills argue that political theory must start by taking into account what humans are like now. The principles of non-ideal theory that they develop are closely tied to actual human motivations and psychologies. And Rawls is somewhere in the middle, tying his two principles to an optimistic view of human nature.

If political theory parallels moral theory in this way, then we have a new kind of ideal/non-ideal distinction within political philosophy. Just as we can have multiple voluntarist constraints involved in our moral theory, so political theory may benefit from multiple voluntarist constraints. Just as we start with ideal moral theory (using a physical voluntarist constraint) and then filter out obligations at each succeeding level of impossibility, so we might start out with ideal political theory and then filter out obligations as we get more and more non-ideal. In Chapter 1, I argued that ideal theory is necessary for providing us a goal to transition to. Viewing ideal theory as using a weaker voluntarist constraint, and non-ideal theory as using stronger versions, provides us with another view of the ideal/non-ideal relationship. Of course the parallels between moral and political theory are not perfect. Because political theory involves many more actors, problems of transition play a much more significant role (this is borne out in the literature on ideal and non-ideal political theory). And perhaps the voluntarist constraint plays different roles in political and moral philosophy. Still, those political theorists who do use it (such as Wiens and Estlund) should accept multiple voluntarist constraints into their theories.

2.5 The relationship to ideal and non-ideal theory

As I have defined ideal theory throughout this dissertation, it in its broadest sense tells us the ideal—it tells us the best version of something. Non-ideal theory tells us what to do when we aren't in that ideal situation. In this case unyielding moral theory (whatever its content) uses a physical voluntarist constraint and is our ideal. As we apply stronger voluntarist constraints, we begin doing non-ideal theory. It's not

ideal that it's psychologically or motivationally impossible for us to live up to unyielding moral theory; we need non-ideal theory to tell us our obligations then.

There are several ways in which the existing literature on ideal and non-ideal theory can shed more light on the theory I have been developing. First, I will better situate this ideal/non-ideal distinction by comparing it to the one Rawls makes, the original distinction in this body of literature. After that, I will discuss a competing notion of feasibility from the ideal and non-ideal theory literature, the one Holly Lawford-Smith offers. I will suggest that the way I have conceived of possibility doesn't draw the artificial distinctions between kinds of feasibility that her conception does. And finally, I will close this chapter with some thoughts on the nature of the ideal/non-ideal distinction itself. The discussion so far shows us that in at least some cases, there is no bright line between ideal and non-ideal; the distinction can be better thought of as a continuum. Thus, while the ideal/non-ideal distinction I have been developing can benefit from considering the existing literature, it stands to be of benefit it as well.

Rawlsian ideal and non-ideal theory

"Ideal" and "non-ideal," like a lot of Rawlsian jargon, have particular meanings for Rawls, and there may be resistance to using them outside of the context of justice. We might worry that expanding the uses of the terms will be confusing or dilute the meaning; we might worry that different uses will have little or nothing in common.

These concerns shouldn't prevent us from considering whether we can expand the uses of these terms. The literature on ideal and non-ideal theory is still in its early days, and it's too early to know whether multiple ideal and non-ideal distinctions will be confusing or helpful. Reconfiguring the jargon might become necessary later on, but there's an important family resemblance—a commitment to discovering some kind of ideal and to figuring out a corresponding set of non-ideal goals, obligations, or strategies. That may turn out to be a limited family resemblance—we may draw the distinction in different ways, the ideal and non-ideal components may be related differently, the obligations to transition may be different. But whether or not the family resemblance helps us to make progress on these questions is, again, a question for later. For now, answering this question of how to apply ideal and non-ideal theory will help us to get a start on answering the theoretical questions later.

In this particular case, however, the Rawlsian distinction can help us out. Remember that for Rawls, ideal theory is distinguished from non-ideal theory in two ways: the ideal has full compliance and favorable background conditions (*Theory* 215). In the non-ideal world, people choose not to fully comply with both principles of justice, and the background conditions of the society are such that the principles cannot be fully implemented. In Chapter 1, I wrote that we can distinguish these parts of non-ideal theory according to whether whether people are or aren't blameworthy for failing to comply with ideal principles of justice.

The way that I've divided up ideal and non-ideal morality doesn't depend on a distinction between background conditions and compliance. But the distinction Rawls

makes has a role to play here. It can help us figure out how to transition to the ideal—our remedy for something we are blameworthy for may be different from our remedy for something that isn't our fault. If it's my own laziness that makes me unable to wake up on a Saturday morning, I need a different remedy than if I have a sleep disorder. It can also help us figure out how to respond to a particular violation of ideal morality. Anger may be the appropriate response if someone is blameworthy for putting a requirement of ideal morality out of reach for herself. If she is not blameworthy, compassion may be the appropriate response; anger is likely to be inappropriate.

Rawls's distinction between compliance and background conditions is helpful in working with the structure of morality that I've outlined, but it's not necessary for understanding what that structure is. The distinctions between the different kinds of possibility that are closer to or further from the ideal and the distinction between compliance and background conditions are, if not exactly orthogonal, at least not dependent on each other. So while there's a family resemblance between Rawls's ideal/non-ideal distinction and mine, they aren't the same distinction. But as we continue to develop ideal and non-ideal theories, about justice as well as about other subjects, perhaps this kind of family resemblance will help us in our analysis. At any rate, both Rawls's ideal theory and the ideal theory I've developed here share a focus on the goal we are trying to reach; both non-ideal theories help tell us how to get there.

Binary and scalar feasibility

I have suggested some ways that the literature on ideal and non-ideal theory can help us understand how ideal and non-ideal moral theory is supposed to work: the attention in the literature to transition helps us see how moral theory can be demanding enough, and Rawls's ideal/non-ideal distinction can help us understand the different paths to the ideal. But putting the distinction to use in the way I have can also give us a way to critique some of the literature. For one thing, it can help us understand feasibility more generally. While feasibility has been thought to be a binary notion (either something is feasible or it isn't), Lawford-Smith has argued that it is part binary, part scalar.[47] But I think applying the notion of feasibility to ideal and non-ideal moral theory shows us that we ought to think of it as entirely scalar.

In Lawford-Smith's terms, binary feasibility is settled by the "hard constraints," what is logically, conceptually, metaphysically, and nomologically impossible (Lawford-Smith 252). This is a weak sense of feasibility. It's where Lawford-Smith locates the voluntarist constraint, since if we allow stronger senses of possibility to determine our obligations, "they'll rule out oughts that shouldn't be ruled out" (254). Scalar feasibility is determined by "soft constraints"—economic, institutional, and cultural constraints, among others (255). A goal that violates the soft constraints isn't *in*feasible, just *less* feasible. When we start talking about scalar feasibility, we are talking about things that are more or less *likely* to happen (Lawford-Smith 255). And we can choose to pursue less feasible outcomes; we just have to keep

[47] For a discussion of feasibility in the literature, see Gilabert and Lawford-Smith.

in mind that we may have to make sacrifices in order to make those less feasible outcomes actual.

But once we see that feasibility is scalar to some extent, it's not clear why we should ever see it as binary. The preceding discussion about ideal and non-ideal moral theory can help us to see why it's hard to draw a line between binary and scalar feasibility, why the constraints of so-called binary feasibility are not all of a piece, and why keeping binary feasibility begs the question against certain ideal theories.

First, where exactly is the line between binary and scalar? Where does something go from being impossible to just infeasible? Lawford-Smith wants scalar possibility to start with something like physical possibility. But some things that are physically possible for me can change, and some can't. It is currently impossible for me to lift 100 pounds, but I could start going to the gym and make it possible. But I could probably never lift half a ton, even with the best trainers in the world. Somewhere in there, Lawford-Smith wants to draw a line between binary and scalar feasibility, between hard and soft constraints. But where would that line go? That it's difficult to tell whether it's a hard or soft constraint suggests that there isn't really a clear-cut distinction between binary and scalar feasibility after all.

Even if we could draw this line, it's not clear what unites the constraints Lawford-Smith thinks determine binary feasibility. Even if a nomological impossibility has the same zero probability as a logical impossibility, there are important differences between the two. Think of the uses of idealized scientific models. Given the laws of nature of our world, it's impossible to slide down a

frictionless plane. But it can be worth considering frictionless planes for modeling purposes when doing physics. The nomological impossibility of frictionless planes helps us to get a handle on what is actually nomologically possible for us. Outcomes that are merely nomologically, as opposed to metaphysically or logically, impossible are closer to our actual world, and they may help us more in thinking about our actual world. This suggests that binary constraints are not all of a piece: the thicker the binary constraint, the more useful it is for us to think about.

Finally, leaving hard constraint-violating theories out of our set of possible unyielding theories begs the question against unyielding theorists who wouldn't make this move: in particular, Cohen, who thinks that fundamental truths about morality are true regardless of any information about humans at all.[48] We get "rules of regulation" for humans by plugging facts about humans into the ultimate principles of morality, but the ultimate principles of justice, the sources of these rules, are unconstrained by any facts at all. So while the voluntarist constraint rules out certain rules of regulation, it does not rule out any ultimate principles of justice or morality.

For Cohen, then, there can't be any clear line between hard and soft constraints, between binary and scalar feasibility. Hard and soft constraints operate in the same way on our options. An option that is ruled out by some quirk of human psychology has the same status as one that's ruled out by a law of nature—both of these options are infeasible rules of regulation, but neither of these cases tells us anything about what the ultimate principles of morality are like. There isn't anything

[48] See esp. 250-254.

special or distinct about hard-constraint-violating options that rules them out of our set of options.

Lawford-Smith has given us very good reasons to think that we should see feasibility partly as a scalar concept—something can be more or less feasible depending on the constraints that apply to it. Considering the nature of feasibility in ideal and non-ideal morality suggests that it is wholly a scalar concept. It's hard to distinguish binary from scalar feasibility; getting rid of binary feasibility means we don't have to treat different kinds of hard constraints as all of a piece; and we no longer beg the question against Cohen. We should abandon binary notions of feasibility and think of feasibility as solely a scalar concept.

The nature of the ideal/non-ideal distinction

Seeing feasibility as wholly scalar can help us to make progress on the nature of the distinction between ideal and non-ideal theory. One of the debates about this distinction is whether these two kinds of theory are sharply distinct from each other or whether they are instead continuous with each other. Proponents of ideal theory have tended to talk about it as though it is sharply distinct from non-ideal theory.

Colin Farrelly has a contrasting view. For Farrelly, theories that abstract more are more ideal; theories that abstract less are more non-ideal (846). But there is no point at which abstracting more is sharply distinguished from abstracting less, where ideal theory is distinguishable from non-ideal theory. Instead, we get a continuum:

Cohen's theory of justice is more ideal than Rawls's is, and Rawls's is in turn more ideal than Carens's is.[49]

Here are two reasons for thinking that the ideal/non-ideal distinction I have drawn is continuous, not sharp. First, the things that separate ideal from non-ideal theory are continuous with each other. We may not be able to draw a clear distinction between physical possibility and psychological possibility and individual psychological possibility and motivational possibility, and that's fine. It may not always be clear on what conception of possibility a given act is possible. What is clear is that things that are less possible are more ideal, and things that are more possible are less ideal.

Second, the relationships between the different levels of theory are the same. The thing that distinguishes ideal theory (with a physical voluntarist constraint) from non-ideal theory (with a psychological voluntarist constraint) is that ideal theory uses a weaker sense of possibility, physical rather than psychological. The thing that separates non-ideal theory (with a psychological voluntarist constraint) from non-ideal theory (with a motivational voluntarist constraint) is that one uses a weaker sense of possibility than the other. Any "bright line" that could be drawn between ideal and non-ideal theory can also be drawn between different levels of non-ideal theory.

And finally, I have been referring to ideal theory as the theory that uses a physical voluntarist constraint. That isn't an arbitrary choice. We are operating under

[49] Farrelly is critical of ideal theory—he believes that too much abstraction leads us to ignore important facts about justice which would affect the shape of our theories (848-56). But this critical project is separable from the descriptive project of putting the ideal/non-ideal distinction in continuous, rather than discrete, terms.

the assumption that some version of the voluntarist constraint is true. Because of this constraint, and because we're trying to come up with morality for human beings, it makes sense to restrict our discussions to what is possible for human beings, although we will want to set our ideal theory at the outer limit of what's possible for humans.

But although this choice isn't arbitrary, it also didn't have to be made this way. Remember the other types of possibility from the beginning of the chapter—logical, metaphysical, nomological. It would have been possible to declare any of these the kind of possibility that's important for determining the ideal. Perhaps the outer bound of our obligations should be determined not by what we're physically capable of but by what's possible given the rules of logic. Cohen might hold such an ultra-idealized view.[50]

There are certainly problems with this kind of view. It seems to respect the letter of the voluntarist constraint but not its spirit; it doesn't seem fair for morality to hold us to obligations that are merely logically possible for us. But whether or not such a view is a good one, the existence of Cohen-style views gives us another reason to think of this form of the ideal/non-ideal distinction as continuous rather than sharp. Just as the different versions of non-ideal theory are continuous with each other and with the ideal, so we could continue further out, to even more ideal versions of ideal theory.

Of course, that doesn't mean that all distinctions between ideal and non-ideal are continuous. In some cases, there may be a sharp distinction that's worth drawing.

---

[50] See esp. ch. 6.

Think about Rawls's ideal theory, which requires full compliance with the principles of justice. There's a clear, binary distinction between full compliance and partial compliance—either you have full compliance or you don't. And there's nothing more ideal than full compliance—it isn't possible to have fuller-than-full compliance. So we should conceive of some ideal theory as continuous with its counterpart non-ideal theory, but that doesn't tell us about all non-ideal theory.

This suggests that the nature of ideal theory may be more diffuse than we might have thought. Rawlsian ideal and non-ideal theory is an ideal and non-ideal theory *of justice.* But we could have tried to port it over to morality with the same structure more or less intact—we could have said that a situation is ideal from the point of morality when we have full compliance with the rules of morality and when we have favorable conditions. But it may be more helpful to think about ideal and non-ideal morality in other ways, such as using the framework I've offered. We don't have to try to force ideal and non-ideal theory into the mold of Rawls. There are other ways we can analyze the distinction between ideal and non-ideal.

How this resolves the two tensions

The early sections were devoted to laying out two tensions in moral theory: first, the tension between unyielding, yielding, and moderate moral theory; and second, the tension between different interpretations of the voluntarist constraint. I then argued that we should adopt ideal and non-ideal moral theory that uses multiple versions of the voluntarist constraint.

It should be fairly obvious how this would resolve the second tension, between different voluntarist constraints. We don't have to decide between voluntarist constraints: we can have them all. If we are trying to figure out the outer limits of our obligations, as Estlund is in thinking about justice, then we should use a thin voluntarist constraint. If we are trying to figure out what we should actually do in a particular set of circumstances, one in which there are psychological or motivational constraints on what we can do, then we should make the voluntarist constraint thicker. In general, I pointed out, different types of possibility are relevant to different situations. The same thing is true in moral theory. The tension between different voluntarist constraints disappears when we see that different constraints are relevant to different things we want moral theory to do.

We can resolve the first tension as well when we see that the three different views of moral theory represent three different views about the proper interpretation of the voluntarist constraint. Unyielding theories, such as act utilitarianism, are unyielding because they operate with a thin voluntarist constraint. The act utilitarian is not constrained by facts about what people are motivationally or psychologically like—whatever we can physically do determines our obligations. Yielding theories are lax because they operate with a very thick voluntarist constraint. We can know our moral obligations only after we know what our personal projects are—in other words, what our motivations are. If a putative moral obligation conflicts with our motivations, we are simply not bound by it. And then moderate moral theories are somewhere in

the middle. Flanagan is concerned to show how morality is psychologically possible for us, while not yielding to our motivations.

That means that the resolution to this first tension parallels the resolution to the second tension. Unyielding, moderate, and yielding versions of moral theory all have their place. Rather than arguing that some version of moral theory is better than another, we should instead see these kinds of moral theory as occupying different spaces on the ideal/non-ideal continuum. Unyielding moral theory represents the ideal—it represents what we would be responsible for doing if we lacked any psychological or motivational limits. It tells us the fullest extent of our moral obligations. But since we live in a non-ideal world in which those limits exist, we need corresponding non-ideal theories. Moderate moral theory gives us a moderate target to aim for in the non-ideal world—it tells us our obligations given our psychological limitations but ignoring our motivational limitations. And the most non-ideal—the most yielding—moral theory takes into account all of our limitations. In doing so, it is best able to give us concrete action guidance but at the cost of ignoring some of our obligations. Whatever ideal moral theory turns out to be right—utilitarianism, deontology, or something else—it will need non-ideal counterparts.

This gives us an answer to the criticisms leveled against each type of moral theory. Each type is vulnerable to certain criticisms only if it is supposed to be the whole of moral theory, not if it isn't supposed to do all the work of morality. Unyielding theories, such as act utilitarianism, are said to be too demanding. But that criticism fails if an unyielding theory is only meant to provide an ideal to reach for,

not immediate action guidance. Yielding theories, such as Williams's, are said to be too lax. But that criticism fails if a yielding theory isn't supposed to tell us all of our obligations. This also helps us to fix moderate moral theory. Above, I noted how Flanagan's views were supposed to represent a compromise between the demandingness of utilitarianism and the laxity of Williams. I argued that this compromise does not work. Moderate moral theory is not a satisfactory answer to the critics of either extreme. It makes sense, however, as one non-ideal component to our ideal moral theory. If we recognize that there are obligations that are motivationally and psychologically impossible, but physically possible, then we can see that there are more ideal versions of morality we should strive for.

Chapter 2, in part, is currently being prepared for submission for publication (as "Ideal Theory and 'Ought Implies Can'"). The dissertation author was the primary investigator and author of this paper.

Chapter Three

Beneficence and Partial Compliance

We have moral duties of *beneficence* to help others who are in need. Some people will fail to comply with these duties. To what extent are compliers required to pick up the slack of non-compliers? If we do not pick up others' slack, then needs go unmet. In the case of global poverty, that means that some people will suffer and die. When we think about it that way, it seems that our duty of beneficence should increase in situations of partial compliance, so that we are required to pick up others' slack. On the other hand, an increase in that duty would ask a lot of us. And more than that, it seems *unfair* to saddle those of us who are doing the right thing with additional duties *because* others are failing to meet their duties. Our punishment for doing our duty by taking on the burdens of compliance is more burdens of compliance.

This is an issue of ideal and non-ideal theory. One way philosophers distinguish ideal from non-ideal theory is that ideal theory assumes full compliance with the rules of morality (e.g., Rawls, *Theory* 215; Valentini, "Conceptual Map" 655-56). Are our duties of beneficence in circumstances of partial compliance limited to what they would be in full compliance, or might we be obligated to do more in non-ideal situations than in the ideal? If our duties do not increase in situations of partial compliance, then non-ideal theory doesn't tell us anything new about our moral duties. If our duties do increase in situations of partial compliance, then while we derive our non-ideal duties from our ideal duties, non-ideal and ideal theory will be distinct.

In this chapter, I take on this issue of our duties in situations of partial compliance. I begin with consequentialism, which has paid a lot of attention to beneficence. Consequentialists disagree about what our duty of beneficence consists in. Peter Singer argues for an extremely demanding duty of beneficence that increases when others don't do their duty. Liam Murphy, on the other hand, argues that there is no difference between our duty in non-ideal circumstances of partial compliance and ideal circumstances of full compliance—we are never obligated to do more than we would have to if everyone were complying. But Murphy's view has implausible consequences: in particular, it seems not to require increased beneficence even in cases of easy rescue. Both of these consequentialist views present us with difficulties: either we have an extremely demanding view of beneficence, on Singer's view, or an implausible one, on Murphy's.

But those are not the only two options. Consider a moral theory that has historically had less to say about beneficence: deontology. Most deontologists accept at least one of these two claims: we have other duties besides beneficence, and at least some of our duties (including beneficence) are imperfect duties. These seem to pose a problem for the idea that our duty of beneficence might increase in situations of partial compliance. If we have other duties besides beneficence, then there are competing moral considerations that restrict the demands beneficence can make on us. And if our duty of beneficence is imperfect—if it is just a duty to do something sometimes—then we may exercise discretion about when and how to be beneficent. If we may exercise discretion, then we don't have to help whenever we can. Doing more to aid others

when need increases would then seem to be supererogatory, not obligatory. An increase in our duty in situations of partial compliance would seem to be more appropriate for consequentialist views. On the other hand, a deontology that is sensitive to the distinction between situations of full and partial compliance could respect the powerful intuition which underlies Singer's view, that it is not enough to do our fair share if the needy continue to suffer.

Deontology can do this. To show how, I look at two deontological accounts of beneficence, Herman's Kantian account and Ross's intuitionism. In both cases, the most sensible interpretation of these accounts allows duties of beneficence to increase in situations of partial compliance, regardless of the distance between us and the recipients of our aid.

These two deontological approaches to beneficence have significant advantages. They can retain the attractions of deontology, such as some room for partiality towards ourselves and our associates. Deontological views also avoid the problems Singer and Murphy face. They don't have to sign up for the implausible consequences of Murphy's view. Unlike Singer's view, they come with natural limits on our duty of beneficence. Deontology thus holds out the possibility of doing a better job of accommodating central beliefs about the nature and limits of beneficence than these two consequentialist views can.

3.1 Consequentialism

Singer

Singer's argument for a demanding duty of beneficence appeals to two simple principles:

1. "Suffering and death from lack of food, shelter, and medical care are bad."
2. "If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it" ("Famine," 231).

It's obvious that we can prevent at least some suffering or death without sacrificing anything of comparable moral importance. My afternoon latte is not comparably morally important to the life I could save with that $4. So I should always give up my latte and donate that money instead. If this argument is sound, our duty of beneficence requires us to give to the level of marginal utility—the level at which we would do more harm to ourselves than we would do good to a needy person (Singer, "Famine" 241).[51]

Singer supports (2) with a famous thought experiment: "If I am walking past a shallow pond and see a child drowning in it, I ought to wade in and pull the child out. This will mean getting my clothes muddy, but this is insignificant, while the death of the child would presumably be a very bad thing" ("Famine" 231). Most of us, Singer thinks, will agree that we should rescue the child, even at this small cost to ourselves. There's just no way that getting my clothes muddy, even ruining them, could ever be comparable to the badness of the death of an innocent child. Singer argues that the case of the distant needy is in all morally relevant respects just like the case of the

[51] Singer offers a more moderate version of this argument as well but says he can see no good reason for affirming the moderate rather than the strong version ("Famine" 241).

drowning child. Even though the distant needy are at a distance to us, we can help them just as easily as we can pull the child out of the pond. So if we are obligated to pull the child out of the pond, we should similarly sacrifice anything that is not of comparable moral importance to saving the lives of the distant needy ("Famine" 232).

It's not surprising that a utilitarian would make this argument. Singer thinks that we are obligated to maximize good consequences (for him, happiness). But which action maximizes good consequences depends on the circumstances, including what others do or fail to do. In nearly every case, it would maximize utility for compliers to give more in situations of partial compliance. It doesn't matter whether others are complying with their duty or not—all that matters is what good you can do. If we accept premises (1) and (2), then we must do all we can to prevent something bad from happening. If others do not work to prevent suffering, then there is more suffering that we are obligated to prevent. Standard consequentialism appears to require this distinction between our duty in ideal and non-ideal worlds: it appears that our duty will increase in the non-ideal world. On Singer's view, our duty of beneficence increases in situations of partial compliance or in any other situation of increased need.

Murphy

A common criticism of Singer's view is that it is too demanding, since it potentially requires compliers to give until they're nearly as badly off as those they're helping. Murphy argues that the demandingness of Singer's view is not the real

problem. I won't go into those arguments here, interesting though they are.[52] Instead, I will focus on the other half of Murphy's view, where he lays out what he believes is the real problem with Singer's position on beneficence: that it is *unfair*. Singer and Murphy agree on the extent of our duty in the ideal, in a situation of full compliance: we should, they agree, divide the duty up in whatever way maximizes good consequences. But they diverge on what happens in situations of partial compliance. Where, as we saw, Singer thinks our duty increases in situations of partial compliance, Murphy argues that we don't have any further duties in these situations. The demands of beneficence in situations of full compliance set a ceiling on its demands in situations of partial compliance.

Murphy thinks of beneficence as a collective duty: it is a duty that belongs to the collective of agents who are in a position to help. We together aim to benefit others as much as possible (Murphy 75). We incur burdens when we comply with the duty: we must give up money and other resources. Because beneficence requires this redistribution of goods, we ought to divide up the burdens fairly among the members of the collective (Murphy 89). Once you know your fair share, you cannot be obligated to do more than that.[53] This would be unfair. It would mean treating the members of the collective who aren't doing their fair share as if they were unable to comply, when in fact they're just slacking (Murphy 116). So Murphy sees no difference in our duty of beneficence between situations of full and partial compliance

[52] For these arguments, see chs. 1-4.
[53] Our burdens may be *different* in cases of partial compliance, but they cannot be *greater.* So, if others are failing to comply with the duty of beneficence, maybe I should switch the organization to which I donate my money, but I cannot be obligated to donate *more* (Murphy, 90).

(that is, between ideal and non-ideal situations). Because Singer's view requires you to pick up the slack of noncompliers, Murphy thinks, it makes fundamentally unfair demands.

Perhaps these two claims, that beneficence is a collective duty and that requiring you to pick up the slack is unfair, are separable. I don't think Murphy makes good arguments for seeing beneficence as a collective duty. He thinks beneficence is collective because we must work together, but there are many non-collective duties we may require others' help to carry out, such as the duty to care for our aging parents. Conversely, we can sometimes act beneficently on our own, such as by donating money. Distinctively collective duties, such as the duty to engage in a fair negotiation, seem to require collective action (Schapiro 333). Beneficence does not.

But whether beneficence is an individual or collective duty, there does seem to be something wrong with asking compliers to pick up noncompliers' slack. Often, this wrongness is expressed in terms of overdemandingness. Murphy wants to resist overdemandingness worries in favor of unfairness worries. Maybe there's something to this. Even if I don't have to do much to pick up your slack, it does seem like there's something unfair about that. But while Murphy has that intuition going for him, his view of beneficence is vulnerable to objections.

Objections to Murphy

Murphy's position seems to have some strange and unpalatable implications. Consider a two-person rescue case.[54] Andrew and Bethany are near a pond, and no other adults are around. Two children are drowning. Andrew saves one of the children, but Bethany keeps walking. Is Andrew required to save the second child? Murphy must say no, as he himself admits (132). Andrew and Bethany are engaged in a collective project of beneficence, so Andrew is only required to do his fair share. It would be good for him to save the second child, but if it requires him to incur any additional burden, it cannot be obligatory.

But this seems absurd. Surely, absent some compelling considerations, Andrew is obligated to save the second child. It can't possibly be permissible for him to let a child drown just because Bethany didn't do her job. Andrew would be a moral monster.

In response to this objection, Murphy claims that most people will, in fact, go beyond their fair share in easy-rescue cases (132).[55] As for those who don't, if we are upset with them, it "is based not so much on a sense that the agent acted terribly wrongly but on a sense that his emotional indifference to the victim's plight shows him to have an appalling character" (Murphy 133). Saving the extra people is supererogatory; not doing so is appalling. But it's hard to figure out what could explain the appalling character of the agent except that he has failed in his duty and

[54] See Singer, Schroeder, Cullity, and Rachels for examples of these cases (Singer, *Life* 144-46; Schroeder 12; Cullity 74-76; Rachels 162-63). Arneson offers a slightly more complicated version (35-39).

[55] He does admit that if the additional rescue costs nothing, then we are obligated to do it, since what we care about is the burdens on the compliers: but obviously this will rarely be the case in the standard situation of beneficence, since rescuing an additional person will cost us money, at the very least (Murphy 128).

acted wrongly. The agent's character is appalling because he failed to do something obligatory, not because he failed to do something supererogatory.

So Murphy badly misses the mark in the two-person rescue case. But then think about fairness. If we require Andrew to pick up Bethany's slack, that seems unfair, even if it's not demanding. But one of the problems we face in doing non-ideal theory is that we are likely to be unable to get rid of all unfairness. Unlike in the ideal, where we might expect our duties to divide fairly and all needs to be met, unfairness will persist when some people are not doing what they ought. Either the compliers will bear the brunt of the unfairness (because they have to do more than they should have to), or the needy will (because their needs will go unmet). In the two-person rescue case, Andrew is being treated unfairly if he must pick up Bethany's slack, but this small unfairness drops out of our moral calculus almost entirely when we think about the second child's needs. Murphy wants to maintain that the unfairness of making the compliers bear extra burdens is a distinct kind of unfairness, more "central" because it is unfairness in the way we design our moral principles, not in their knock-on effects (90-91). But even if this is true, the two-person case shows that Murphy gives his "central" sense of unfairness far too much weight (see also Newey 48 and Cullity 77). Unfairness may become an issue if an increase in the duty of beneficence requires us to do much more, but then the problem arises from the demandingness of our burdens, not from their being unfairly imposed. Small unfairnesses are an unfortunate but unavoidable side effect of our non-ideal world.

This unfairness is mitigated when we see that this doesn't mean letting noncompliers off the hook entirely. Murphy thinks that if compliers are required to save additional people—if they are required to pick up noncompliers' slack—that we are letting the noncompliers off the hook and failing to treat them as agents (116). But just because we do their job for them, that doesn't mean that we've let them off the hook entirely. They're blameworthy for their failures to comply. We might hold them responsible for causing us to take on greater moral obligations, even if they can no longer be responsible for the solutions to those problems (Miller 237). If we fail to fulfill the additional obligations we incur because of partial compliance, we are not responsible in the same way as the noncompliers are (Miller 244). We can blame them, resent them, express our disgust at them, and so on. This is something we do all the time when we're engaged in collective projects. The college student whose friends slack off on the group project might stay up until 2 am to get it done, but this will affect how she thinks of them in the future. The wife whose husband refuses to do laundry may do it because it needs to get done, but she may become angry or ask him to make up for it by doing the dishes more often. Just because others' duties fall to us doesn't mean we let them off the hook completely.

While unfairness might be mitigated in this way, it may still remain troubling. But we will have unfairness somewhere in non-ideal theory. Murphy maintains fairness in the two-person case at the cost of deeply held intuitions about easy rescue, and I think his answer there is extremely implausible. Murphy's concerns about fairness should not get the weight he gives them. But that doesn't mean there's

nothing regrettable about increasing the burdens on compliers. As we will see, though, a less demanding deontological view of beneficence has less trouble with unfairness.

So we get a dilemma. Murphy's version of beneficence seems open to obvious counterexamples. If there are two children drowning, and Bethany doesn't do her fair share, it seems bizarre and terrible to suggest that Andrew is not morally required to rescue the other child. On the other hand, Singer's version of beneficence opens us up to nearly limitless demands.[56] If the world we live in is anything like our current world, we must reduce ourselves to near-poverty just to do our duty. Singer's view, unlike Murphy's, can respect the intuition that we must, at least sometimes, pick up others' slack. It just appears to go too far in how much slack it requires us to pick up.

3.2 Deontology

Some people are motivated by the extreme demandingness of views such as Singer's to search for a deontological alternative. Deontology defies easy description, but I will begin by focusing on two claims that consequentialists typically make:

1. There is only one duty, which is the duty to promote the good—that is, the duty of beneficence.
2. All duties are perfect duties, meaning that it is always wrong to fail to perform them, in any context.

Deontologists deny at least one of these claims. They are what makes Singer's view so demanding. We must do as much good as we can—and when we fail to do as much

[56] Cullity provides theoretical backing for the intuition that Singer's view is too demanding. See esp. Chs. 7 and 8.

good as we can, we are failing to comply with our (one and only) perfect duty.[57] Deontologists also reject (1) and (2), while maintaining that we have moral duties.

There are two questions I want to ask now. First, can deontologists do *as well as* consequentialists? Are there any plausible deontological moral theories that respect the intuition underlying Singer's position, that we must at least sometimes pick up others' slack? Second, can deontologists do *better than* the consequentialist views I have discussed? Can they avoid the demandingness complaints associated with Singer's view?

They can. To show how, I will examine two deontological views of beneficence. First is Ross's intuitionism. Ross rejects claim 1—on his moral theory, there are a plurality of duties that together determine what we ought to do in a given situation. Because Ross relies on the weights of the competing duties to determine what we ought to do, his theory makes sense of the intuition that our duty of beneficence is weightier when need is greater. Ross may also reject claim 2, but for a more sustained discussion of perfect and imperfect duties I turn to Herman, whose Kantian account of beneficence, I will argue, requires us to do more when the need is greater, whether the needy are nearby or complete strangers to us.

If I am right, then two deontological accounts of beneficence can make room for the intuition that makes Singer's account more plausible than Murphy's. They both require us to pick up some slack when others fail to comply. But because these views recognize multiple duties, and because at least some of these duties are imperfect, they

[57] Note that Murphy may not be a consequentialist on this definition: he does not appear to accept the first claim, because he accepts other moral principles besides beneficence (75). But then his view, as I have argued, is implausible.

have an important strength that Singer's view does not: they are not nearly as demanding. Both of these views can make room for other kinds of moral obligation that many of us think are important: duties to ourselves and to our associates, duties to never perform certain sorts of actions and to respect certain rights. Moreover, both of these views can give us space for discretion about how to act morally or whether to act for moral reasons at all. This means we are permitted to act on our own desires and projects at least some of the time. Thus deontological views offer a good way to capture two compelling ideas about morality: against Singer, they place limits on the demands moral theory can make of us; against Murphy, they recognize that we may sometimes have to pick up others' slack.

The denial of claim 1: Ross and pluralism

Ross develops an intuitionist form of deontology on which we have multiple irreducible duties: so he denies consequentialist claim (1). Ross names seven duties, the first six of which are fidelity, reparation, gratitude, justice, self-improvement, and non-maleficence (21).[58] The seventh is beneficence, which rests "on the mere fact that there are other beings in the world whose condition we can make better in respect of virtue, or intelligence, or pleasure" (Ross 21). Beneficence is the only "consequentialist" duty, in the sense that it is the only duty explicitly concerned with maximizing overall good consequences. Because there are other duties, the theory as a whole is not consequentialist.

[58] Although Ross also tries reducing some of these duties to others and considers the possibility that there could be additional duties (26).

These seven duties are all *prima facie* moral duties (Ross 19).[59] A prima facie duty is a duty we must act on unless it is outweighed by some other prima facie duty. What we must actually do in a particular situation is determined by whatever duty has the greatest weight—this duty becomes our duty *sans phrase*. We determine this by looking at the features of the situation we are in (Ross 19). Perhaps I could either give you money when you're starving (beneficence) or use that money to pay back my friend for buying lunch (gratitude). Given the circumstances, I intuit that beneficence is weightier than gratitude, so that is my duty sans phrase here.

Ross likens the process of figuring out our duty sans phrase to understanding the laws of motion: "*Qua* subject to the force of gravitation towards some other body, each body tends to move in a particular direction with a particular velocity; but its actual movement depends on *all* the forces to which it is subject" (28-9). We cannot break a promise to one person in order to deliver one more unit of good to another (Ross 34). In this case, the duty of fidelity exerts a stronger force than the duty of beneficence does. But, Ross implies, if there were a greater disparity between the weights of the duties, we might be justified in breaking the promise (35). Perhaps we would be justified if we could deliver a thousand more units of good by breaking an unimportant promise, or a million. But where beneficence actually comes to outweigh fidelity is left to our intuition.

This is not to say that Ross thinks all the prima facie duties are equally weighty. What Ross identifies as the "perfect duties" (fidelity, reparation, and

[59] Ross notes that "prima facie" isn't exactly the right term, since it implies duties that disappear under certain circumstances, but uses the term anyway (20).

gratitude) are more stringent than the rest (41). But these are not perfect duties in the sense of consequentialist claim (2) (Ross 2). Ross explicitly rejects Kant's view of perfect duties that "admit of no exception in favor of imperfect duties" (18). For Ross, the label "perfect" means that the duty is presumptively weightier, not that it is always weightier. If self-improvement is weighty enough in some case, then we can act on that duty rather than on fidelity. (Perhaps I have made some nearly meaningless promise to you that will require me to miss the SATs or an important job interview.) Our duty sans phrase can sometimes be an imperfect duty, even if a perfect duty is also applicable to the situation.[60]

Ross and partial compliance

The question we have been considering is how to act in situations of partial compliance with the duty of beneficence. What does Ross's view have to say here? In a situation of full compliance, by definition, everyone complies with beneficence when it is her duty sans phrase. Because others are doing their duty, there is less overall need. When everyone else is complying with beneficence, it takes up less of our moral space. Other concerns will take priority. Each of us will certainly sometimes need to help others; beneficence will sometimes win out over other duties. But there will be many other times where it does not. We will have to take it into consideration every time, but it will often lose out against our other prima facie duties.

[60] It's possible that this means that Ross accepts consequentialist claim 2, that all of our duties are perfect duties. Clearly none of the prima facie duties is a perfect duty. But if all of our duties sans phrase are perfect, then there is a sense in which all of our duties are perfect. If we must always act on whatever our duty sans phrase is, then our duty sans phrase is a perfect duty. Prima facie duties are imperfect; sans phrase duties are perfect. See Hill's interpretation of Ross ("Beneficencde" 14). But Ross is at any rate not a consequentialist, since he rejects claim 1.

But in our own non-ideal world, where compliance failures are massive, many more people will need our help than in the case of full compliance. So our duty of beneficence will become more stringent. The additional need increases the gravitational pull beneficence has, since each of us is in a position to do so much more good for which there is an increased need. Compare the above example of fidelity and beneficence. It's clear that the reason beneficence comes to outweigh fidelity is that the good consequences of acting beneficently eventually outweigh the importance of promise-keeping. One extra unit of good is not enough to outweigh the importance of promise-keeping, but a much more significant increase in good might be. In situations of partial compliance, beneficence takes on this extra importance.

But beneficence is not the only duty that bears on this question. The fact that we are in a situation of partial compliance *itself* adds to the overall judgment that we must do more to help the distant needy. For Ross, justice is a duty to ensure that goods are distributed proportionally to the merit of the people concerned (21). The distant needy deserve aid that they aren't getting because of compliance failures, and so they are on the receiving end of a great injustice. We may not be able to do anything about those who unjustly benefit from their own failure to comply, but we can restore some balance for those who are unjustly harmed by compliance failures. So justice, along with beneficence, seems to require additional aid in situations of partial compliance.

But of course there are countervailing justice considerations. Think about Murphy's claim that asking compliers to pick up the slack of noncompliers is fundamentally unfair. I mentioned above that I am skeptical of the view of

beneficence as a collective duty that would give rise to these concerns about fairness. But Ross gives us a different kind of reason to care about fairness. If compliers must give up more of their resources to pick up the slack, then goods are distributed out of proportion to merit—because noncompliers don't do their duty, and we do, we have less than we should have, and they have more. This is true whether beneficence is a collective or individual duty. The duty not to steal is surely an individual duty, but if you steal from me, then our goods are distributed out of proportion to our merit. So Ross can take fairness into account without saying anything about whether beneficence is a collective duty.

Perhaps in true intuitionist fashion, it's not immediately clear where this leaves us. I have tried to adduce some factors that are relevant to situations of partial compliance. It seems relevant that need is much greater when many people fail to comply with their duties of beneficence. It seems relevant that, because the distant needy are not getting what they're owed, they are victims of injustice. These factors push us in the direction of thinking that beneficence (or, perhaps, the combination of beneficence and justice) becomes more stringent relative to other obligations in situations of partial compliance. On the other hand, it seems relevant that an increase in our duties in situations of partial compliance results in our having less than we are due. And perhaps other duties are relevant as well.

Although our duties are left up to intuition, there's a good case for thinking that overall our prima facie duty of beneficence is weightier in situations of partial compliance. As need becomes greater and greater, we are in a position to do vast

amounts of good. And while the injustice of making sacrifices we shouldn't have to make is an important consideration, it may be outweighed by the similar, but worse, injustice done to the distant needy by making them bear the burdens of others' noncompliance. The distant needy are suffering, so our duty of beneficence applies. They are also victims of a great injustice, so our duty of justice applies. The injustice done to us by having to pick up others' slack, especially when we are otherwise in possession of at least as many goods as we deserve, pales in comparison. But on an intuitionist theory, our best guess at what our duty really is will be determined at least in part by the specific circumstances in which we are choosing how much to give.

Plural vs. single duties

When setting out the contrast between consequentialism and deontology, I said that the question for deontological theories is whether they can reject the two consequentialist claims (that there is only one duty; that all of our duties are perfect) while respecting the intuition that we must sometimes pick up others' slack. Ross's deontology provides an example of how this can work. Situations of partial compliance, I have argued, provide the additional stringency we need to justify adding weight to our prima facie duty of beneficence. This means that our duty sans phrase will be a duty of beneficence more often in situations of partial compliance than in situations of full compliance. At the same time, concerns of justice—concerns about taking on extra burdens caused by others' noncompliance—are not entirely absent. So Murphy's concerns are alive in a Ross-style account of beneficence and partial

compliance as well.[61] We will have to rely on our intuitions about whether the injustice done to us when we pick up the slack of noncompliers outweighs the injustice done to the needy who suffer when their needs go unfulfilled. In some cases, it may. But we can accept a plurality of duties and still think our duty of beneficence increases in situations of partial compliance.

So Ross's theory bests Murphy's. Think about the case that Murphy handled so poorly: the case in which two children are drowning, Bethany refuses to save one, and so it seems like Andrew is required to save both. Murphy thought Andrew had no obligation *whatsoever* to save the second child. Surely Ross, who wants to save our common moral beliefs, can deliver the result that the stringency of Andrew's duty here outweighs most other prima facie duties he might have.

But Ross can also beat out Singer. We were looking for a moral theory that could not only accommodate the intuition behind Singer's view but do so without being unreasonably demanding. Ross's pluralism about duties can recognize the complexity of our moral lives and so take care of the demandingness objection. One duty Ross recognizes is the duty of self-improvement, the duty we all have to "improve our own condition in respect of virtue or intelligence" (21). Duties to self will not always win out over duties to others, but there will probably be some circumstances in which our duty of self-improvement will be weightier and thus allow us to prioritize our needs over the needs of others, even if meeting their needs brings about better consequences overall.

[61] In fact, Murphy mentions offhand that Ross's theory is likely to be extremely demanding, because it contains a consequentialist view of beneficence (10n4). But since it also includes *other* principles, it is less demanding than full-blown consequentialism is.

More speculatively, we might apply an intuitionistic method not just to the question of *which* prima facie moral duty is our duty sans phrase in a given situation but also to *whether* we have a duty sans phrase at all. There may be times when we *could* help someone, or increase our own virtue or intelligence, or work for justice (that is, there might be some prima facie duties that are relevant to our situation), but we might intuit that it is permissible for us to not act on any of these duties at all. Consequentialists will naturally be suspicious of this result. And there are some passages in Ross that contravene this idea: consider his offhand remark that we may *always* be in a situation in which at least one prima facie duty is incumbent on us, in which case we *always* have a duty sans phrase (19).[62] But Ross's larger project is to find a moral theory that fits with common-sense moral thinking (17-19). An intuitionism that has us use our intuition not just about *what* duty we have but *whether* we have a duty at all fits with a common-sense view of morality's demands. Many people have the intuition that at some point we are allowed to stop acting morally: that we have some space for discretion about how to live our lives, some "free time" to pursue our desires and interests. Deontologists who adopt Ross's framework have a way to deny consequentialist claim (1), while nevertheless accepting a rise in beneficence in cases of partial compliance, that is nonetheless not as demanding as Singer's view.

The denial of claim 2: Herman and imperfect duties

Imperfect duties

[62] Thanks to Charlotte Newey on this. See also Hill, who takes Ross's view to imply "that it is one's actual duty to promote others' happiness on every occasion when one can and other duties are absent" ("Beneficence" 14).

I said that two claims distinguish consequentialism from nonconsequentialism: (1) that beneficence is the only moral duty, and (2) that all moral duties are perfect duties. In looking at Ross, we saw how a moral theory with a plurality of duties can give us an increase in our duty of beneficence in situations of partial compliance. But Ross has an idiosyncratic view of what perfect and imperfect duties are. In fact, he criticizes Kant for employing perfect duties in his account (Ross 18). And yet the distinction between perfect and imperfect duties seems to be an important part of the moral architecture for many deontologists. So it would be good to see whether a view that makes the standard perfect/imperfect distinction can reach this same result. I will argue that it can. Here, I take as representative Herman's Kantian view of our duty of beneficence. Herman accepts the distinction between perfect and imperfect duties. Her way of spelling out what imperfect duties are requires us to increase our beneficence in situations of partial compliance. But because Herman's account is Kantian, it still lacks the extreme demandingness of Singer's view. This means that Herman's view, like Ross's, meets the desiderata I laid out earlier: it is nonconsequentialist, in virtue of rejecting claims (1) and (2), it can do *as well as* consequentialism on the question of partial compliance, and it can do *better than* consequentialism on the issue of demandingness.

Before getting into the specifics of Herman's view, we should look at the distinction between perfect and imperfect duties. In my discussion of consequentialist claims (1) and (2), I defined perfect duties as duties which we must always perform, regardless of context. "Don't lie," "don't murder," and "don't cheat" are perfect

duties—it would *always,* without exception, be wrong to lie, murder, or cheat.[63] In contrast, imperfect duties come with exceptions. Specifying what this means is tricky. Rainbolt lists eight different possible ways to understand the distinction (233). Stohr defines imperfect duties as duties to adopt maxims (50). Schroeder thinks they are duties that give us some latitude, but he points out difficulties with specifying exactly what that means (1-5). I think Hill puts it well: imperfect duties require us to "take to heart certain principles, not that we act in certain ways" ("Kant" 57).[64] These are duties because they create some kind of moral requirement, but it is not a requirement in every context. Beneficence is a standard example of an imperfect duty: many deontologists who think we have the duty to help others do not think that we have to help everyone whom we could possibly help. We can decide whom to help and when to help. But we must take seriously our obligation to take others into account when deciding what to do—we must take to heart the maxim of helping others. So one way of understanding imperfect duties is as duties to adopt certain policies—duties to make certain things, such as the need of others, relevant in our deliberations about what to do.[65] Making others' needs relevant, or taking them to heart, does not, however, require us always to act on those needs.

---

[63] Of course, actual perfect duties may be more complicated than simply "never steal."

[64] Stohr says something similar, although she conceives of beneficence as having a complicated structure that combines perfect and imperfect duty (58-67).

[65] Hill moots the possibility that Kant holds a different view of imperfect duties—that while we can choose *which* imperfect duty to act on, we must act on *some* imperfect duty when we have the opportunity to ("Kant" 58). We can choose to act beneficently or to improve ourselves, but we can't permissibly choose neither. Sometimes Herman seems to suggest this as the way to understand imperfect duties, although her considered view is not completely clear ("Scope" 238-41). But Hill rejects this interpretation on textual grounds, and it seems overly rigoristic as well ("Kant" 58-60).

What makes imperfect duties imperfect? Guyer suggests that it is because it would be impossible for beneficence to be an imperfect duty because we cannot possibly help everyone (194). But that can't be it. Consequentialists know that we cannot possibly help everyone, but they argue that the perfect duty of beneficence requires us to help others whenever we can. For deontologists, though, even if I *could* possibly help everyone, beneficence would still be an imperfect duty. There must be some other reason that imperfect duties are imperfect. Morality, the nonconsequentialist might say, has to give us some room for discretion, especially when its requirements are potentially very demanding. Exercising this discretion can be an important part of my moral life. Even if I am capable of giving everyone the help he needs, morality should not *ask* me to help everyone. Instead, when and how and whom to help is up to my discretion (up to a point—I must at least help *sometimes*).

Consider beneficence, standardly thought of as an imperfect duty.[66] It was easy to explain why a perfect duty of beneficence would increase in situations of partial compliance. If you always have to act beneficently, regardless of the context, then it doesn't matter whether you happen to be in a context where lots of people are failing to comply with their duty. But if beneficence is an imperfect duty, then a rise in

[66] Perhaps we have a limited perfect duty of beneficence in cases of easy rescue. We might think that we are obligated to perform every easy rescue we can, but that other beneficent acts are the fulfillment of imperfect duties. On the other hand, what if you're in a situation in which you are confronted with a series of easy rescues that would prevent you from doing anything else with your life (see Timmerman)? (We might think that this describes our actual world, in which you can easily save many lives by giving your money to charity.) Maybe these all count as easy rescues, in which case the perfect duty of beneficence eats up our entire lives; or maybe at some point the burdens these "easy" rescues place on us is substantial enough that "easy" rescue no longer passes the second test, limiting our perfect duty to truly easy rescues. So if there is a perfect duty of beneficence, there are questions about its scope. But set these aside to focus on the imperfect duty of beneficence.

situations of partial compliance seems less obvious. The imperfect duties are duties to do something sometimes; they are duties that come with latitude in deciding when and how to carry them out. If we are already doing something sometimes, how can it matter for our duty whether others are failing to do something sometimes, or in fact whether need has increased at all? If we have latitude in deciding how to carry out our duty of beneficence, shouldn't that latitude preclude a rise in the duty based on the circumstances we're in? On some views, perhaps it does. But on Herman's view, which I turn to now, our imperfect duties increase when need increases.

Herman on beneficence

Herman's rationale for the duty of beneficence is the familiar Kantian argument for all of our moral duties: is it a universalizable maxim? We cannot rationally will a world in which nobody helps us when we need help. This means that we cannot universalize the maxim "never help anyone," so we cannot rationally will a world in which we never help anyone (Herman, "Scope" 232-33). This gives us a duty to help at least some people sometimes—an imperfect duty of beneficence.

Because we are rational agents, we take on goals and projects that make us happy or contribute to our wellbeing (Herman, "Scope" 241). Sometimes, we need aid with these projects.[67] This means that the duty of beneficence must extend beyond easy rescue. If all we could rationally will were help in order to keep us alive, then the duty of beneficence would simply be a duty to keep others alive. But because we need

[67] For complicated reasons, Kant thinks that we can only affect others' happiness, not their rational agency (see for example "Metaphysics" 517-18). But we must try to bring about their happiness out of respect for their rational agency.

help with the projects that are connected to our happiness, the duty of beneficence must be directed at others' happiness.

This is at least related to the consequentialist understanding of what beneficence is (an understanding that Ross shares, since beneficence is for him a consequentialist principle within a deontological framework). For Singer, beneficence is directed at maximizing happiness, understood as pleasure. For Herman, beneficence is also directed at happiness, although not at *maximizing* happiness (since, as we will see below, you are permitted to favor your loved ones over strangers), and at a happiness that is connected to rational agency. Herman and Singer may vary somewhat in the extensions of their duties of beneficence, but at base they share a concern for the happiness of others.

But Herman introduces an additional complication into her analysis of beneficence. Some of our imperfect duties to ourselves, she argues, *necessarily* precede our duties to others: "unless one is willing and to some degree able to enjoy life, one cannot appreciate and so correctly evaluate the range of human concerns. One will not make wise judgments about either one's own needs as an agent, or about the happiness of others" (Herman, "Scope" 242). Because our duty of beneficence is directed at increasing the happiness of rational agents, we need to know something about what happiness is for rational agents—about how to determine what a good project or goal is, about how to gather the resources necessary for completing that project or goal, about what it's like to complete or adjust or dispose of a project or goal. In order to know what a good life is like for a human, that is, we need to know

what a good life is like for ourselves. Thus we may not forgo, for example, our education in favor of acting beneficently, because to do so will harm our long-term ability to help others (Herman, "Scope" 244). This does *not* mean that we must have the freedom to live very luxurious lives—Herman thinks this is impermissible ("Scope" 255). We must simply be able to enjoy life *to some degree*, by making *some* choices about our projects or getting *some* latitude to decide what our lives should look like.[68]

So we can favor ourselves, and Herman also gives us a justification for favoring those with whom we stand in some kind of special relationship, even on top of whatever associative duties we have to them ("Scope" 253). We are usually better at furthering the projects of those close to us. We know more about what those projects are and what will make them work than we do about the projects of a complete stranger. We can tailor our aid much more precisely than we can when we're giving it to someone we've never met. So we are better able to assist those we know, or perhaps those who are in some kind of community with us, than we are to assist strangers. We can more efficiently attend to their needs as rational agents than we can to the needs of those we don't know. To Herman, this fits well with "everyday morality," which is "inherently local" ("Scope" 230).

[68] A different way to get to a similar conclusion is Cullity's argument in Ch. 8. One reason to save others' lives, he argues, is the goods they get from living, including friendships and personal achievements. So we have moral reasons to save people living non-altruistically-focused lives. But if it is permissible for those others to lead non-altruistically-focused lives, then it must also be permissible for us to live non-altruistically-focused lives (while also leaving some room for beneficence). This is close to Herman's conclusion but without the epistemic premise that we need our own non-altruistically-focused lives in order to understand why this is desirable.

But even though we can give special consideration to those we know, we have a duty of beneficence to every rational agent, including those who are strangers to us. For the most part, Herman thinks, our obligations of beneficence to the distant needy take a special shape. The distant needy are needy because their own societies have failed to provide for them, which means that our obligations of beneficence are "inherited obligations" that we inherit thanks to the failure of someone else (Herman, "Scope" 249). For example, food aid to the distant needy must not hinder a local society from providing its own food (Herman, "Scope" 250-51). So the way we carry out our duty of beneficence to the distant needy may be different from the way we carry out our duty to those we know. But it's still more than just a duty of rescue: it's a duty to provide whatever is necessary for "adequate social and economic functioning, as these are understood locally," since it's a duty to aid the development of others' happiness (Herman, "Scope" 251).

So, to sum up finally: our duty of beneficence is an imperfect duty which is directed at increasing the happiness of others and which can only be limited by other imperfect and perfect duties. While the ways in which these duties limit each other cannot always be known in advance, some duties to self necessarily precede duties to others, because those duties to self are necessary in order to successfully perform acts of beneficence. We are also sometimes permitted to favor our associates (friends, family, fellow citizens) over the distant needy.

Not every account of beneficence as an imperfect duty will have all of these features. We might disagree with Herman that rational agency is really the reason we

have a duty of beneficence, or we might disagree that we need experience with rational agency in order to be able to carry out our duty of beneficence. But Herman's account is, at least, one of the major attempts to elucidate Kantian thinking on beneficence, and Kantian deontology is, at least, one of the major strands of deontology that employs a perfect/imperfect distinction. How does it handle situations of partial compliance?

Partial compliance and the nearby needy

I will start with the easy cases for Herman. It's clear that our imperfect duty of beneficence to associates increases in situations of partial compliance. Our duty to the distant needy is a little trickier, and I will look at that last. But every kind of beneficence, I will argue, increases in situations of increased need, including situations of partial compliance.

In the first case, that of those we know (friends, family members), our duty of beneficence may be quite demanding, even in normal circumstances. Although some duties to ourselves will always come first, most of us will find ourselves involved intimately with others as we create relationships with them. And the more people we have relationships with, the more we are in a position to act beneficently by increasing their happiness (Herman, "Scope" 247). This means that we can't know what our duty of beneficence to those we know looks like until we know what our relationships with those people will be (Herman, "Scope" 247). Our duty of beneficence increases, whether in situations of full or partial compliance, as we come to know more people (Herman, "Scope" 245). These people are at the fore of our moral concern, so we have

an extensive imperfect duty of beneficence to them, even in situations of full compliance.

In the next case, we see that our duties to fellow citizens increase in certain circumstances: “If, living in a just society, one happens to be the person in front of whom large numbers of people trip and fall, then one is unlucky, and large demands are indeed made on one’s time and resources. There can be no moral guarantee that one will get to live the life one wants” (Herman, “Scope” 249). Not everyone is in a circumstance where lots of need exists, but if that’s the circumstance you are in, you are required to do something about it. We can look backward from the case of fellow citizens to the case of friends and family. If our obligation to meet the needs of our fellow citizens increases if they are falling through the institutional cracks, then surely our obligation to act beneficently towards those we know increases as well.

But why do these imperfect duties increase? Think about how we fleshed out what an imperfect duty is: it is a duty to adopt a policy of helping sometimes, a duty to take certain principles to heart. What it is to adopt a policy is to make beneficence relevant in our deliberations about what to do. If I have a policy of helping others in general, but all of a sudden many people are tripping and falling around me, and yet I am not acting any differently, do I really have a policy of helping at all? If I only ever have a chance to help one person, and I help that one person, then we can fairly say I might have had a policy of acting beneficently. But if I can help a thousand people, and I only help one person once, then it becomes much harder to say that I have a policy of helping, rather than attributing my beneficence in that one case to an

accident or a whim. Herman does not think that we can necessarily say "in advance or in the abstract" where to draw the line between helping ourselves and helping others—there is no single quantity of beneficent action that will show that we have adopted a policy of acting beneficently ("Scope" 243). Still, if we rarely or never act beneficently in the face of significant opportunity, that is good evidence for our not having that policy. If there are more opportunities to help—especially if I am in a position of being able to be especially helpful, either because of a special relationship or simple proximity—the amount of aid I render should increase. We care not just about acting on the right maxim but on the success of our actions (Herman, *Practice* 98). If we have adopted a maxim of helping others, and need is going unmet, we renew our efforts.

Adopting a policy of acting beneficently thus means making the possibility of beneficence relevant in our deliberations about what to do. We must consider the ways our money could be used to help others when we are deciding how to spend it. And uses of our money that were permissible when everyone was complying with the duty of beneficence will become impermissible as need grows, because spending our money on luxuries for ourselves in those circumstances will indicate that we did not carefully consider the dire need we could have been addressing with that money. Other imperfect duties will work the same way. Friendship permits you to ignore some of your friend's minor concerns, but a true friend cannot ignore her friend's desperate pleas for help. The person who ignores her friend in desperate circumstances isn't a friend at all—she has failed to give her friendship the weight it deserves.

On Herman's view, we must balance imperfect duties against each other. Our duty to promote our own perfection balances against our duty to promote the happiness of others, but some level of perfection in rational agency (and thus, some striving for happiness) is a necessary precursor to understanding how to promote others' happiness. As Herman herself notes, this will be a complex balancing, and "we may not be able to say in advance or in the abstract where the line is to be drawn between what we require for ourselves and what can permissibly be made available for others" ("Scope" 243). But whatever the balance is, it will come with a substantial duty of beneficence.

But perhaps we can say a little more than that about where the line is to be drawn. If needs, whether of our family, friends, or the distant needy, are systematically going unmet, perhaps we ought to shift the balance of duties away from ourselves and toward others. If the best way for me to understand the value of rational agency is to take on a project of devoting myself to raising the finest foie gras in the country, but others are starving, maybe I should develop my rational agency a little less well by doing something that's a little less resource-intensive. Herman notes that most of us choose projects that will be able to survive interruption, in case others (particularly those we know and our fellow citizens) suddenly impose significant burdens on us ("Scope" 253). What I am suggesting goes a little further. If there is great need in the world, we should not just choose projects that can survive interruption. We should also choose projects (where we can) that will free up more resources for others. If there is great need, we should stack the deck in favor of our duty of beneficence and against

our duty to ourselves. This will of course not mean that we never have concern for ourselves—part of Herman's account is that such concern is morally required. And it will not mean that we abandon partiality toward those we know—our duties to the distant needy must fit with the relational duties we have (Herman, "Scope" 253). But as we develop our rational agency, we attune our capacities of concern for others (Herman, "Scope" 245). Surely this ought to lead us to consider how our projects can fit not just with our concern for those we know, but also for those needy we may never meet.

But the added concern for beneficence doesn't rule out supererogation, which is also an important part of the moral architecture for deontologists. We have not done away with latitude altogether. Although we must do more to act beneficently in situations where need is great than when it is small, beneficence is still an imperfect duty. We can still trade it off against other imperfect duties and even decline to act on any imperfect duty at all. So supererogatory action is still possible when we give beneficence outsize importance in our deliberations and so always or almost always put beneficence ahead of our own desires.

Partial compliance and the distant needy

Now we come to the distant needy, the paradigm case of beneficence. If our duty of beneficence to associates increases in situations of partial compliance, it might seem straightforward to think our duty to the distant needy increases as well. Oddly, Herman disagrees: "When we are not in a position to exercise judgment, because need is at a distance, or the needy are strangers to us, or private charity is inappropriate,

public institutions can do the work of beneficence for us, and that part of our general duty is met by contributing a fair share of support" (Obligatory Ends" 273-74). This looks like Murphy's view. But this can't be right, for several reasons.

Implicit in the idea that "we are not in a position to exercise judgment" is the plausible claim that we can do much more to bring about effective agency in those close to us because we know better what they need. We can usually only give money to the distant needy—to people we know, we can give care, cultural goods, and so on (Herman, "Obligatory Ends" 273). It's true that we can't exercise judgment about the way Oxfam spends our money in the same way that we can exercise judgment about when to give a friend an aspirin, so those near to us sometimes take precedence. But we do know some things. We know that people need food, water, and sanitation in order to survive and become rational agents. We also can make pretty good guesses at some other things agents need, such as education. So we can exercise at least limited judgment about the distant needy: what organizations are doing the most for their happiness and how we can contribute to those organizations.

But even if we couldn't exercise any judgment, why would this lead us to a fair-share view? That would be inconsistent with Herman's own account. She writes that our obligation to the distant needy arises "because, given the obligatory end of the happiness of others, we already have an indeterminate obligation to all persons that bears on their need" (Herman, "Scope" 252). Our duty of beneficence, no matter its target, is derived from our awareness that our happiness depends on the happiness and agency of others. The distant needy are rational agents just like our associates are; we

owe them the same duty of beneficence (even if we can prioritize our family over strangers to some degree). Thus, an imperfect duty of beneficence increases in situations of partial compliance, regardless of the people to whom we owe that duty.

Imperfect vs. perfect duties of beneficence

In my discussion of Ross, I argued that beneficence can be one duty among many and still increase in circumstances of partial compliance. Herman also admits of a plurality of duties, but she adds an additional wrinkle: the distinction between perfect and imperfect duties. If we understand imperfect duties as duties to do something sometimes, we might wonder how a duty to do something sometimes can increase in situations of partial compliance. But when we think more about what doing something sometimes means—if it means making beneficence relevant in our deliberations about what to do—then our duty of beneficence increases as need increases. Adopting a policy of acting beneficently, understood in this way, gives us an increase in our duty of beneficence.

How does Herman do compared to Murphy and Singer? Clearly, she does not give us the implausible answer Murphy does in the two-person rescue case. For one thing, we may have a perfect duty of easy rescue that would require Andrew to rescue the second child. But even if beneficence is only an imperfect duty, beneficence that is truly deliberatively relevant for Andrew will require him to pick up Bethany's slack. This is just like one of the tripping-and-falling cases I mentioned earlier. As we saw then, Herman's view may be quite demanding in cases of "moral misfortune," where

an individual is confronted with many demands that are small individually but over time lead to a significant cumulative burden on that individual ("Mutual Aid" 598).

Despite this, Herman's duty of beneficence is still not as demanding as Singer's is. Beneficence is still mainly an imperfect duty: even in situations of increased need, we have some latitude about whether to act beneficently. Making beneficence deliberatively relevant does not mean it always defeats other considerations. And beneficence must contend with other imperfect duties—our duties to ourselves and to those we know. In some cases, our duties to ourselves will necessarily precede our duties to others. We must first ensure that we have the goods we need in order to develop as agents, and after that we must do the same for people we have some kind of relationship with. So while Herman's view, like Singer's, demands that we pick up other people's slack, it retains the duties to self and associates that are characteristic of deontological views. We can accept that our duty increases in situations of partial compliance without having to be impartial consequentialists. So a view of beneficence that adopts a structure of perfect and imperfect duties can give us a rise in our duties in situations of partial compliance (or other increased need) without giving us the implausible answers Murphy does or the extreme demands Singer does.

Conclusion

The motivation for this chapter was to see whether deontological theories of beneficence could accommodate an intuition that I find powerful: that in situations of partial compliance (or any situation of increased need), it is not enough to do what you

would do if everyone were complying with the duty of beneficence. Your duty of beneficence must increase. I wanted to know whether a moral theory could accommodate that intuition and yet preserve other features of deontology that I find attractive: its plural duties, its distinction between perfect and imperfect duties, its lack of extreme demandingness. I have shown here that it can. Whether deontology is a Ross-style intuitionism or a Kantian account of imperfect duties, it can show us why we are required to pick up others' slack.

Let me be clear about what I have not shown. I have not provided an argument for Ross over Herman or vice versa. My project here was to show that deontology *can* in various guises show why our duty of beneficence increases in situations of partial compliance, not to prove that Ross or Herman has the one true moral theory. I also do not have an answer for those who embrace Singer's horn of the dilemma by accepting an extremely demanding morality.[69] For those people, there is no reason to prefer a deontological account of beneficence over the standard utilitarian one. Finally, consequentialists may still be able to avoid the Singer-Murphy dilemma by finding a consequentialist theory of beneficence that can allow the demands on us to increase in situations of partial compliance without being extremely demanding. Scheffler and Ridge offer theories of this nature. But the discussion in this chapter shows that we shouldn't count deontology out. For those of us who think morality's demands should have limits, but who are troubled by the suffering of others, deontology may turn out to provide the best account of beneficence yet.

[69] Although, again, see Cullity, Chs. 7 and 8.

Chapter Four

Conclusion

In various ways, my dissertation has been an argument for the usefulness of ideal theory. In Chapter One, I argued that ideals can be necessary when their purpose is to deliver sustained social progress over time. Without ideals, we will be trapped in a version of the problem of second best: we will be unable to make sure that our comparative analyses of possible societies don't lead us down a dead-end street. In Chapter Two, I showed one way that ideals can be useful in moral and political theory. One kind of ideal theory, theory that uses a thin version of the voluntarist constraint, is the starting point for figuring out what more non-ideal versions of that theory we can use to guide our actions. And in Chapter Three, I showed that ideal beneficence can be a starting point for figuring out what we must do to help others and how we must do it.

But we also need non-ideal theory. In Chapter One, I agreed with Sen that ideal theory doesn't tell us everything about what to do. We need non-ideal theory to tell us what to do in non-ideal circumstances. Non-ideal theory may also be all that we need when we aren't trying to make big time-consuming social changes. If there is an obvious ill we can remove, non-ideal theory shows us how.

In Chapter Two, I show how we need thicker voluntarist constraints, pegged to psychological and motivational senses of possibility, to tell us what to do in non-ideal situations. If we think the voluntarist constraint only means "ought implies physically can," then moral theories will be unreasonably demanding for those of us (all of us?) whose psychological and motivational limits prevent us from doing everything that's

physically possible for us. And although our non-ideal obligations are derived from our ideal ones, this derivation is not wholly straightforward. In Chapter Two, I rehearse the example of an arachnophobe who is psychologically incapable of preventing a child from being bitten by a spider. That arachnophobe incurs other moral obligations, such as the obligation to visit the child in the hospital. These are not straightforwardly derivable from any obligation the arachnophobe has on ideal theory; we need non-ideal theory to tell us what to do here.

Chapter Three probably has the most explicit argument against the sufficiency of ideal theory. Murphy thinks that ideal theory is sufficient for telling us what to do in non-ideal cases of partial compliance. In the ideal, everyone would comply fully—therefore, we never have an obligation to do anything other than what we must do in the ideal. But Murphy's view can't even handle a two-person rescue case satisfactorily, and he bumps up against the intuition that we must do more when others are doing less. We need non-ideal theory to tell us what to do here.

4.1 The definitions of ideal and non-ideal theory

In the introduction to my dissertation, I talked about controversy over the definitions of "ideal theory" and "non-ideal theory." I said that I would be using an ecumenical definition: *Ideal theory tells us about the best version of something. Non-ideal theory tells us what to do when we aren't in the ideal (whether because we won't or because we can't)*. The ideal and non-ideal theories I discuss in three chapters of my dissertation all come under this definition. The ideally just society exhibits the best version of justice, but we need both transitional and non-transitional non-ideal theory

to tell us how to get there. Ideal moral theory tells us what we would do if we were the best versions of ourselves (lacking our psychological and motivational weaknesses), and non-ideal moral theory tells us what to do as we are. Ideal beneficence would involve full compliance, but non-ideal beneficence does not.

A more substantive definition of ideal theory would not capture all three of these topics. As I said in my introduction, one such definition is that ideal theory means full compliance; that is, everyone complies with the principles (usually, of justice) (Valentini, "Conceptual Map" 655-56). Full compliance plays a major role in Chapter Three. The definition of ideal beneficence Murphy proposes just is your share of beneficence in a situation of full compliance. Full compliance is also a branch of the Rawlsian ideal theory of justice that I defend against Sen's criticisms in Chapter One.

But Chapter Two doesn't really say anything about full compliance. Here, I describe a framework for different sets of moral rules that are true whether or not anyone complies with them. You have multiple sets of obligations that are possible for you (in different senses of the word "possible), no matter whether you fully or partially comply with them. The ideal theory of morality is determined by what you are physically capable of doing, not by what people will comply with or what you in particular will comply with. Non-ideal theory isn't determined by the effect partial compliance has on morality; instead, it's determined by a thicker sense of "can." Full compliance plays no role in constructing this kind of ideal theory. And yet the ideal theory of Chapter Two should be considered part of the ideal-theory family. The

distinction between ideal and non-ideal moral theories of this kind plays the same role as it does in ideal and non-ideal theories of justice: ideal moral theory presents a goal "that we are to achieve if we can" (Rawls, *Theory* 216).

Moreover, thinking about the ideal theories that come under the umbrella of this more ecumenical definition helps us to make progress on some of the theoretical debates about ideal and non-ideal theory. One of these is the debate over whether ideal theory is on a continuum with non-ideal theory. The ideal and non-ideal theories of Chapter Two are clearly continuous with each other. This shows that there are plausibly continuum versions of ideal and non-ideal theory. Ideal theory of justice may not have that relationship to non-ideal theory of justice (either you have full compliance or you don't), but this is why we should talk about ideal and non-ideal theory outside of discussions of justice. We can make progress on our understanding of ideal and non-ideal theory this way.

4.2 The necessity of ideal theory

Another connection between the three chapters is that in all of them ideal theory is used to derive non-ideal theory, to varying degrees. The alternative view is the conception of non-ideal theory as independent that Sen touts and that is popular among other recent critics of ideal theory (Anderson, Mills, and so on). One major theme of my dissertation is that this often will not work. In Chapter One, I talk about the problem of second best and local peaks. Non-ideal theory done independently of ideal theory is vulnerable to the problem of second best, because we will not be able to tell whether the state of affairs that looks better will be good for our long-term

progress. This isn't to say that non-ideal theory practiced independently of ideal theory will *never* work, as I have said. Where there are problems we can remove or failures we can analyze without using ideal theory, it may be easier to do so. We may be able to get rid of Jim Crow laws without having to replace them with ideally racially just laws. We don't need to shoehorn in ideal theory everywhere. But we do need it where we want to make sustained progress over time.

In Chapter Two, ideal theory is the solution to one criticism of very yielding moral theories, such as Bernard Williams's. As Flanagan argues, these moral theories are too lax; they do not take account of what we are psychologically capable of doing (for example, it seems like many of us can put our own projects aside if we recognize that morality tells us to). But these moral theories are not too lax if they are the non-ideal companions to ideal moral theory. We can derive non-ideal theory with a motivational voluntarist constraint from ideal moral theory with a physical voluntarist constraint. This will give us something very much like Williams's moral theory but without its attendant problems.

The relationship of the ideal theory in Chapter Three to non-ideal theory is a little different. Here, we may not actually need ideal theory to determine our obligations. Murphy thinks we do—he thinks all we have to do is whatever we would have been obligated to do in the ideal. The two-person rescue case, however, shows that his approach won't work. This means that we need non-ideal theory of beneficence as well. The deontological theories I discuss can show us how to determine our non-ideal obligations: Ross looks at how weighty beneficence is when

others are not complying, and the imperfect duty Herman argues for will become more deliberatively relevant as need rises. But we may not need to know what we would do in a world of full compliance in order to know what we should do right now, since against Murphy, we don't stop with our "fair share" of beneficence.

But there are ways in which ideal beneficence is helpful. As Herman points out, some of our obligations of beneficence are "inherited obligations" ("Scope" 249-55). We inherit them because some other group who would normally have had those obligations has failed to act on them; we inherit corrupt or resource-poor governments' obligations to feed their citizens. Herman points out that many of these inherited obligations will probably get different treatment than our first-line obligations. If we inherit the obligation to feed another country's citizens, that will probably mean, among other things, trying to build up institutions in that other country so that it can eventually take care of its own. Knowing about ideal beneficence can also help us to hold others responsible. While "fair shares" of beneficence do not matter in the way Murphy thinks they do, we can still hold others responsible for increasing our burdens of beneficence. That might change the way we treat people—we would be licensed, at least, to resent them. But of course this happens with inherited obligations that are both individual and collective, just like I might resent that someone's individual duty to care for her elderly parents falls to me.

So we can see that the relationship between ideal and non-ideal theory varies by context. In the first two chapters of my dissertation, ideal theory is necessary for understanding what we must do in the non-ideal world. But when we get to

beneficence, we don't need to know about the ideal in order to know what we should do. Here, ideal theory is helpful but not necessary. This reflects one of the commitments I discussed in the introduction. One way to make progress on ideal and non-ideal theory is to *do* ideal and non-ideal theory in various contexts. Once we do that, we can see where the various versions of the distinction are similar and where they differ. Comparing these three chapters shows us that we cannot say that ideal theory is always necessary. When we think about the roles ideals can play in normative theory, we must be sensitive to the work we need them to do.

4.3 Moral and political philosophy

Another theme in my dissertation is the connections between moral and political philosophy. As I talk about in Chapter One, recent discussions of ideal theory get their start in Rawls. But both moral and political philosophy prescribe ideals that individuals or societies frequently can't or won't live up to. The connections become more apparent when we use a more ecumenical definition of the ideal/non-ideal distinction. When we can consider ideal and non-ideal theory free of the assumptions we make about ideal theory of justice (such as the full-compliance assumption Rawls makes), we can begin to get a grasp of the ideal/non-ideal distinction as such. When we look at lots of different ideal and non-ideal theories, we get a sense of what they all have in common.

One of the places I make the moral-political connection most explicitly is in Chapter Two. There, I talk about a conversation that runs in parallel in political and moral philosophy, about which voluntarist constraint to use. You need to know which

"can" you should use to derive your own individual obligations, but we need to know which "can" we get our political obligations from as well. I show how we can arrange both moral and political theories on a spectrum depending on which voluntarist constraint they use. While I focus on the upshot of multiple voluntarist constraints for moral philosophy, a similar conclusion applies to political philosophy as well. Here, we see how using this methodology in both subfields at once brings out parallels we might not have noticed otherwise.

This chapter also gives us a different way to arrange political theories in terms of ideal and non-ideal theory (that is, by thinking about which voluntarist constraint each uses) from what we see in Rawls. When we see that we can arrange moral theories in terms of which voluntarist constraints they use, we can then apply the same method to arrange political theories. But whether this is equally useful in political philosophy depends on answering other questions about the parallels between moral and political possibility. If something is motivationally impossible for me, I can take steps to make it become motivationally possible: I can go to therapy or read self-help books. But the question of how we make something politically impossible into something possible is much more fraught, because we are now dealing with large groups of people. Parallels between the moral and the political may tend to break down, but more investigation of the role of the voluntarist constraint in political theory could tell us how far this parallel extends

Another payoff of thinking about moral and political philosophy is in Chapter Three. Global justice sits right on the border of moral and political philosophy. My

focus is on the moral aspects of beneficence, since I concentrate on the question of what we as individuals must do when other individuals aren't doing their duty. But this could have ramifications for political issues of global justice as well. We could have a Murphy-style view of our obligations as nations to other nations—whatever our duty to poorer nations is, we might be engaged in a project of collective beneficence with other wealthy countries, and we might not have to do anything more than our fair share.

I argue that this is false in the individual case, but is it in the global case? The parallels between the moral and the political case may break down again. It's unlikely that someone would look at my charitable giving and think that I've given enough to get him off the hook. I don't give that much, or that publicly. But if some states appear to be doing enough to aid the very poor, other states may abstain, and things will go worse. Think of the United States' frustration at a perceived lack of defense spending by other NATO countries on the belief that the US will pick up their slack). Perhaps the political case is collective in a way that the individual case is not. I argue in Chapter Three that Herman (along with Ross) presents a view of beneficence that is better than Murphy's or Singer's, but interestingly Herman herself thinks that governments have no obligations of beneficence to other governments ("Scope" 252). So here the parallel between the moral and the political case might not be exact, but thinking about the moral case can help us see where the issues might lie in the political case.

A final theme in my dissertation is the distinction between transitional and non-transitional non-ideal theory. As I have defined this distinction, transitional non-ideal theory tells us how to get from our current situation to the ideal one, and non-transitional non-ideal theory tells us what to do in the meantime. I talked about this in Chapter One, where I laid out an interpretation of Rawls's version of the ideal/non-ideal distinction. Although Rawls doesn't explicitly talk about the distinction between transitional and non-transitional non-ideal theory, we can see traces of it in his discussion of the general conception of justice. And issues of transitional and non-transitional non-ideal theory are central to the last part of that chapter, where I discuss how we might actually construct an ideal theory we could use to guide our transition.

This distinction also shows up in Chapter Two, where I talk about the problems associated with moving from a non-ideal to a more ideal moral theory. In that chapter, I am laying out a model for moral theories in general to follow, regardless of which ideal moral theory is true. But using that model for any particular moral theory will mean confronting these questions of transition. If we are utilitarians, questions of transition will be pretty simple. If spending my effort on expanding my motivational capacities will maximize happiness in the long run, then that's what I should do; if spending it on doing what I can, given my present motivations, will maximize happiness, then that's what I should do. But questions of transition are harder for nonconsequentialist theories. Right now, I cannot be motivated to always treat others as ends in themselves, but I know it's physically possible for me to do so.

If treating you as a mere means will get me the money for therapy to overcome my manipulative streak, should I do it? Here the answer is not so clear.

One function of non-transitional non-ideal theory is to set boundaries on transitional ideal theory. In the political case, we might think that there are some human rights that cannot be infringed in order to attain massive economic growth that will benefit everyone in the future. In the moral case, the injunction against treating people as means might mean that I cannot use you to get money to avoid using others later. But how non-transitional non-ideal theory sets limits on transitional, and when, is a big question that deserves more exploration.

4.4 The future of ideal theory

In this vein, I want to close by pointing to some directions for future research. Chapter One leads us naturally from early to later Rawls. In *Political Liberalism,* the later Rawls is was concerned to show how we can have justice in the face of reasonable disagreement. I pursue a similar project in Chapter One. How is it possible to do ideal theory when we disagree about what our ideals should be? In that chapter, I argued that we should borrow Cass Sunstein's method of incompletely theorized agreements. This has overlap with Rawls's methodology in *Political Liberalism,* where we find the overlap among reasonable comprehensive doctrines. But this overlap opens up a new line of inquiry.

Political liberalism is supposed to be an ideal theory of justice which improves on Rawls's attempt in *A Theory of Justice* by taking account of reasonable disagreement. But we have reasonable disagreement because of the burdens of

judgment. Biases and gaps in information are what lead to our inability to agree on what the ideal would be. So political liberalism has non-ideal theory at its heart. As long as we carry the burdens of judgment, we are in non-ideal circumstances. On the other hand, Rawls's turn to political liberalism represents recognition of an important fact, that one of the cornerstones of modern liberalism is respect for the views of others and an acceptance of diversity in thought and action. One task for future research is understanding how the ideal/non-ideal distinction interacts with political liberalism. We must investigate how we can preserve diversity of thought in the ideal, because this is one of the major advances of political liberalism. At the same time, we must consider how much diversity of thought comes out of non-ideal circumstances because of the burdens of judgment.

I have shown in this dissertation that there are many different versions of the distinction between ideal and non-ideal theory. One direction for future research is to think about other versions of the ideal/non-ideal distinction in moral theory. Chapter Two describes one of those distinctions, according to the interpretation of the voluntarist constraint. But there must be others. For instance, what role does ideal virtue play in moral theory? Are there distinct virtues of non-ideal morality, or how do ideal virtues guide our non-ideal actions? Wolf's paper on moral saints, in which she questions whether we would actually want to be or know people with all the virtues to the highest degree, might be a good reference point here. So there's room for much more work on ideal and non-ideal distinctions within moral theory.

Chapter Three beckons to a number of interesting topics for future research. One question is how much Murphy's and Singer's positions on beneficence are affected by whether each views beneficence as a collective duty. I suspect that this is not a negligible difference. It's because Murphy thinks that we hold the duty of beneficence collectively that we can think about how to divide it fairly among ourselves. Meanwhile, Singer just thinks we are each responsible for doing the most good we can do. It doesn't matter morally why there's need; all that matters is what we can do. I'm skeptical of Murphy's reasons for thinking that beneficence is a collective duty (although others frequently assume it as well), and showing that it's not a collective duty would bolster the argument that we must do more when others are not doing their duties.[70]

Another way to go is to apply the conclusions from Chapter Three to another problem, climate change. Our duty to prevent or slow climate change is like our duty of beneficence: it is a duty to make progress on an enormous problem, and many people are not complying with it. This will lead to disastrous consequences. If we have a duty to pick up others' slack where beneficence is concerned, surely we also have a duty to pick up others' environmental slack, if for no other reason than that the magnitude of the problem is so great. To show that, though, we need a more thorough investigation into our moral duties to preserve a habitable climate. It might also require showing that duties against climate change, like beneficence, are individual rather than collective duties. And there are almost certainly other duties that have the

[70] See Schroeder and Cullity, among others.

form of beneficence and duties against climate change. One of the main upshots of Chapter Three was that deontology is better equipped to handle beneficence than consequentialism is. Future research might show whether this is true of these other issues as well.

And finally, future research might investigate one of the main themes of my dissertation, the overall usefulness of ideal theory. I have shown some areas where ideal theory is necessary, against non-ideal theorists. But I have also conceded that ideal theory may not be necessary in absolutely every circumstance. Perhaps when we need to make only small changes, or get rid of obvious problems, we don't need ideal theory. Continuing to discover where our ideals can help us to make philosophical progress must be our ongoing task.

Works Cited

80,000 Hours. The Center for Effective Altruism, 2015. <https://80000hours.org> 24 June 2015.

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

Aristotle. *Nicomachean Ethics.* Ed. Roger Crisp. Cambridge: Cambridge University, 2000.

Arneson, Richard. "Moral Limits on the Demands of Beneficence?" In *The Ethics of Assistance: Morality and the Distant Needy,* ed. Deen Chatterjee. Cambridge: Cambridge University, 2004.

Boot, Martijn. "The Aim of a Theory of Justice." *Social Theory and Practice* 15.7 (2011): 7-21.

Brink, David. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University, 1989.

Chuard, Philippe and Nicholas Southwood. "Epistemic Norms without Voluntary Control." *Noûs* 43.4 (2009): 599-632.

Cohen, GA. *Rescuing Justice and Equality.* Cambridge: Harvard University, 2008.

Cullity, Garrett. *The Moral Demands of Affluence*. Oxford: Oxford University, 2004.

Doris, John. *Lack of Character.* Cambridge: Cambridge University, 2002.

Enoch, David. "The Disorder of Public Reason." *Ethics* 124.1 (2013): 141-76.

Estlund, David. "Human Nature and the Limits (If Any) of Political Philosophy." *Philosophy and Public Affairs* 39.3 (2011): 207-37.

Eyal, Nir and Christopher Woodard. "Varieties of Idealizing about Compliance." American Philosophical Association – Pacific Division. Westin Bayshore Vancouver, Vancouver, BC. 2 April 2015. Colloquium.

Farrelly, Colin. "Justice in Ideal Theory: A Refutation." *Political Studies* 55.4 (2007): 844-64.

Flanagan, Owen. *Varieties of Moral Personality.* Cambridge: Harvard University, 1991.

Gaus, Gerald. *The Order of Public Reason.* Cambridge: Cambridge University Press, 2011.

Gheaus, Anca. “The Feasibility Constraint on the Concept of Justice.” *Philosophical Quarterly* 63.252 (2013): 445-64.

Gilabert, Pablo and Holly Lawford-Smith. “Political Feasibility: A Conceptual Exploration.” *Political Studies* 60.4 (2012): 809-25.

Goodin, Robert. “Political Ideals and Political Practice.” *British Journal of Political Science* 25.1 (1995): 37-56.

Graham, Peter A. “‘Ought’ and Ability.” *Philosophical Review* 120.3 (2011): 337-382.

Guyer, Paul. *Kant.* London: Routledge, 2006.

Herman, Barbara. “Mutual Aid and Respect for Persons.” *Ethics* 94.4 (1984): 577-602.

---. “Obligatory Ends.” In *Moral Literacy.* Cambridge: Harvard University, 2008. 254-75.

---. *The Practice of Moral Judgment.* Cambridge: Harvard University, 1993.

---. “The Scope of Moral Requirement.” *Philosophy and Public Affairs* 30.3 (2002): 227-56.

Hill, Thomas. “Beneficence and Self-Love: A Kantian Perspective.” *Social Philosophy and Policy* 10.1 (1993): 1-23.

---. “Kant on Imperfect Duty and Supererogation.” *Kant-Studien* 62:1-4 (1971): 55-76.

Howard-Snyder, Frances. “Ought Implies Can.” *International Encyclopedia of Ethics.* Ed. Hugh LaFolette. Hoboken: Blackwell Publishing, 2013. 1-9.

Jay, Christopher. “Impossible Obligations Are Not Necessarily Deliberatively Pointless.” *Proceedings of the Aristotelian Society* 113.3 (2013): 381-88.

Kant, Immanuel. “Groundwork of the Metaphysics of Morals.” In *Practical Philosophy,* trans., ed. Mary Gregor. Cambridge: Cambridge University, 1996. 33-58.

---. “The Metaphysics of Morals.” In *Practical Philosophy,* trans., ed. Mary Gregor. Cambridge: Cambridge University, 1996. 37-108.

---. “On a Supposed Right to Lie from Philanthropy.” In *Practical Philosophy,* trans., ed. Mary Gregor. Cambridge: Cambridge University, 1996. 605-15.

Kekes, John. “‘Ought Implies Can’ and Two Kinds of Morality.” *Philosophical Quarterly* 34.137 (1984): 459-67.

King, Alex. “Actions That We Ought, but Can’t.” *Ratio* 27.3 (2014): 316-27.

Korsgaard, Christine. “The Right to Lie: Kant on Dealing with Evil.” *Philosophy and Public Affairs* 15.4 (1986): 325-49.

Kratzer, Angelika. “What ‘Must’ and ‘Can’ Must and Can Mean.” *Linguistics and Philosophy* 1 (1977): 337-55.

Lawford-Smith, Holly. “Understanding Political Feasibility.” *Journal of Political Philosophy* 21.3 (2013): 243-59.

Lipsey, RG and Kelvin Lancaster. “The General Theory of Second Best.” *Review of Economic Studies* 24.1 (1956-57): 11-32.

Mason, Andrew. “Just Constraints.” *British Journal of Political Science* 34.2 (2004): 251-68.

Miller, David. “Taking Up the Slack? Responsibility and Justice in Situations of Partial Compliance.” In *Responsibility and Distributive Justice,* ed. Carl Knight and Zofia Stemplowska. Oxford: Oxford University, 2011. 230-45.

Mills, Charles. “‘Ideal Theory’ as Ideology.” *Hypatia* 20.3 (2005): 165-84.

Mizrahi, Moti. “‘Ought’ Does Not Imply ‘Can,’” *Philosophical Frontiers* 4.1 (2009): 19-35.

Murphy, Liam. *Moral Demands in Nonideal Theory.* Oxford: Oxford University, 2000.

Pereboom, Derk. “Kant’s Transcendental Arguments.” *Stanford Encyclopedia of Philosophy.* Stanford University, 2013.

Newey, Charlotte. *Fairness, Moral Demandingness, and Global Poverty.* Diss. University of Reading, 2015.

Quong, Jonathan. *Liberalism without Perfection.* Oxford: Oxford University Press, 2011.

Rachels, James. “Killing and Starving to Death.” *Philosophy* 54.208 (1979), 159-171.

Räikkä, Juha. “The Feasibility Condition in Political Theory.” *Journal of Political Philosophy* 6.1 (1998): 27-40.

Railton, Peter. “Alienation, Consequentialism, and the Demands of Morality.” *Philosophy and Public Affairs* 13.2 (1984): 134-71.

Rainbolt, George. “Perfect and Imperfect Obligations.” *Philosophical Studies* 98 (2000): 233-56.

Rawls, John. *The Law of Peoples.* Cambridge: Harvard University Press, 1999.

---. *Political Liberalism.* New York: Columbia University Press. 2005.

---. *A Theory of Justice*. Rev. ed. Cambridge: Belknap Press of Harvard University Press, 1999.

Ridge, Michael. "Fairness and Non-Compliance." In *Partiality and Impartiality*, ed. Brian Feltham and John Cottingham. Oxford: Oxford University, 2010. 194-222.

Robeyns, Ingrid. "Ideal Theory in Theory and Practice." *Social Theory and Practice* 34.3 (2008): 341-62.

Rohlf, Michael. "Immanuel Kant." *Stanford Encyclopedia of Philosophy.* Stanford University, 2010.

Ross, WD. *The Right and the Good.* Oxford: Clarendon, 1965.

Schapiro, Tamar. "Compliance, Complicity, and the Nature of Non-Ideal Conditions." *Journal of Philosophy* 100.7 (2003): 329-55.

Scheffler, Samuel. *Human Morality.* New York: Oxford, 1992.

---. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions.* Oxford: Clarendon, 1992.

Scheueler, George. *Desire: Its Role in Practical Reason and the Explanation of Action.* Cambridge: Bradford, 1995.

Schroeder, Andrew. "Imperfect Duties, Group Obligations, and Beneficence." *Journal of Moral Philosophy* 11.5 (2014): 557-84.

Sen, Amartya. *The Idea of Justice.* Cambridge: Belknap Press of Harvard University Press. 2009.

Simmons, A. John. "Ideal and Non-Ideal Theory." *Philosophy and Public Affairs* 38.1 (2010): 5-36.

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1.3 (1972): 229-43.

---. *The Life You Can Save.* New York: Random House. 2010.

---. *The Most Good You Can Do.* New Haven: Yale, 2015.

Sinnott-Armstrong, Walter. "Moral Dilemmas and 'Ought and Ought Not.'" *Canadian Journal of Philosophy* 17.1 (1987): 127-39.

---. "'Ought' Conversationally Implies 'Can.'" *Philosophical Review* 93.2 (1984): 249-61.

Stemplowska, Zofia and Adam Swift. “Ideal and Non-Ideal Theory.” *Oxford Handbook of Political Philosophy,* ed. D. Estlund. Oxford: Oxford University Press, 2012.

Stern, Robert. “Does ‘Ought’ Imply ‘Can’? And Did Kant Think It Does?” *Utilitas* 16.1 (2004): 42-61.

---. “Transcendental Arguments.” *Stanford Encyclopedia of Philosophy.* Stanford University, 2015.

Stocker, Michael. “‘Ought’ and ‘Can.’” *Australasian Journal of Philosophy* 49.3 (1971): 303-16.

Stohr, Karen. “Kantian Beneficence and the Problem of Obligatory Aid.” *Journal of Moral Philosophy* 8 (2011): 45-67.

Streumer, Bart. “Does ‘Ought’ Conversationally Implicate ‘Can’?” *European Journal of Philosophy* 11.2 (2003): 219-28.

Sunstein, Cass. “Incompletely Theorized Agreements.” *Harvard Law Review* 108 (1995): 1733-72.

Swanson, Christine. “Satisficing and Perfectionism in Virtue Ethics.” *Satisficing and Maximizing: Moral Theorists on Practical Reason.* Ed. Michael Byron. Cambridge: Cambridge University, 2004. 176-89.

Thomas, Alan. “Sen on Rawls’s ‘Transcendental Institutionalism: An Analysis and Critique.” *European Journal of Political Theory* 13.3 (2014): 241-63.

Timmerman, Travis. “Sometimes There Is Nothing Wrong with Letting a Child Drown.” *Analysis* 75.2 (2015): 204-12.

Valentini, Laura. “Ideal vs. Non-Ideal Theory: A Conceptual Map.” *Philosophy Compass* 7.9 (2012): 654-64.

---. “On the Apparent Paradox of Ideal Theory.” *Journal of Political Philosophy* 17.3 (2009): 332-55.

---. “A Paradigm Shift in Theorizing about Justice?: A Critique of Sen.” Centre for the Study of Social Justice Working Paper (2010), 2-14.

Van Someren Greve, Rob. “‘Ought,’ ‘Can,’ and Fairness.” *Ethical Theory and Moral Practice* 17.5 (2014): 913-22.

Värynen, Pekka. “Ethical Theories and Moral Guidance.” *Utilitas* 18.3 (2006): 291-309.

Vranas, Peter. “I Ought, Therefore I Can.” *Philosophical Studies* 136.2 (2007): 167-216.

Wiens, David. “Motivational Limits on the Demands of Justice.” Version 2.2, August 2014. Unpublished.

---. “Prescribing Institutions without Ideal Theory.” *Journal of Political Philosophy* 20.1 (2012): 45-70.

Williams, Bernard. “A Critique of Utilitarianism.” *Utilitarianism: For and Against.* JJC Smart and Bernard Williams. Cambridge: Cambridge University, 1973.

---. *Moral Luck.* Cambridge: Cambridge University, 1981.

Wolf , Susan. “Moral Saints.” *Journal of Philosophy* 79.8 (1982): 419-39.

Ypi, Lea. *Global Justice and Avant-Garde Political Agency*. Oxford: Oxford University Press, 2011.