

SOCIAL IDEALS AND SOCIAL REFORM:  
POLITICAL PHILOSOPHY FOR A PROGRESSIVE SOCIETY

by

Jacob Barrett

---

Copyright © Jacob Barrett 2020

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2020

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Jacob Barrett titled: Social Ideals and Social Reform

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

*Gerald Gaus*

Gerald F Gaus

Date: May 20, 2020

*Allen Buchanan*

Allen Buchanan

Date: Jun 8, 2020

*Thomas Christiano*

Thomas D Christiano

Date: May 21, 2020

*Connie S. Rosati*

Connie S Rosati

Date: May 21, 2020

*David J Schmidt*

David J Schmidt

Date: May 23, 2020

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

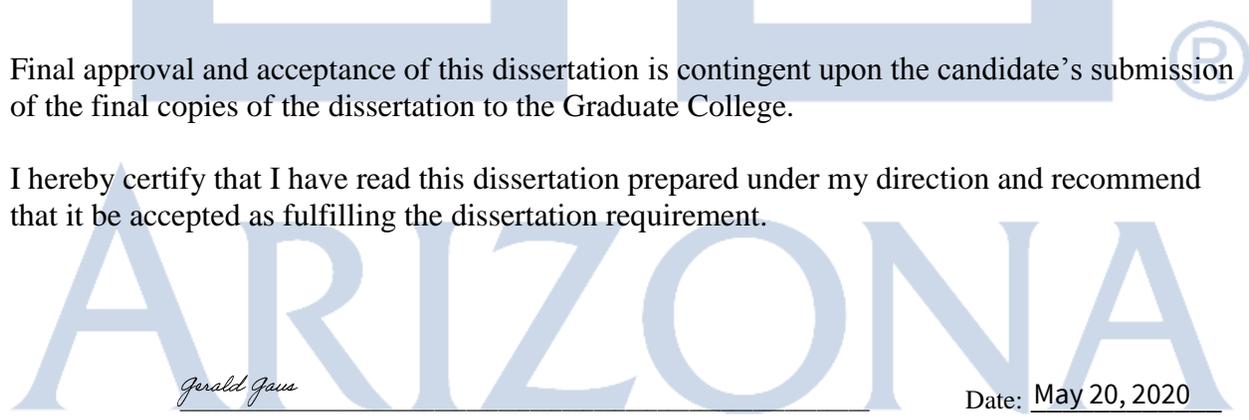
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

*Gerald Gaus*

Gerald F Gaus

Philosophy

Date: May 20, 2020



## ACKNOWLEDGEMENTS

This dissertation has been a long time coming, and I have accrued many debts along the way. I cannot hope to register them all here. So I will just list those that are most salient to me, with apologies to the many others who have helped me throughout this process.

My most obvious debt is to my dissertation chair and advisor, Jerry Gaus, for his endless support and guidance. His influence on my thinking should be obvious to anyone with a passing familiarity with his work, and I have lost track of all the parts of this dissertation that were written expressly in order to follow up on a suggestion of Jerry's or, perhaps more frequently, to respond to an initially devastating objection. Jerry has been remarkably generous with his time and patience throughout my PhD—both during the periods of high productivity in which I made great demands on him to read (often much too long) papers and chapters, and through the lulls where I couldn't seem to get anything onto a page. He is never an easy critic, but is always a fair one, and I would not be the philosopher I am today without him. Nor would this work exist in anything like its present form.

I also owe many thanks to my other committee members, Allen Buchanan, Thomas Christiano, Connie Rosati, and David Schmidtz. Allen has the uncanny ability to provide helpful and encouraging comments on nearly everything I send him in less than 48 hours, and to have already published a paper on just about every topic I raise with him. Tom, who probably remains the least convinced of anyone by the central arguments that follow, proved an incredibly willing and useful interlocutor—there was no better litmus test throughout this process that I might actually be right about something than that I could convince Tom I wasn't obviously wrong. Connie always helped me to work through my ideas and pages with a fine-toothed comb, providing thorough and honest feedback both about what she thought worked and what she thought needed more work; she also went far above and beyond her

formal roles as a committee member or placement director in providing me with timely and spot-on professional advice about everything from the most minute (such as how to write a letter in response to a request for a revise and resubmit) to the most significant (such as how I should spend the next few years of my life). And I cannot count the number of times I have finally come up with an “aha!” solution to something I have been puzzling over, only to realize that its kernel came from a throwaway comment that Dave made to me a few weeks (or sometimes even years) earlier. Much like Jerry, his influence on the shape of this project should be abundantly clear.

In addition to my committee members, I owe thanks to many other friends and colleagues at the University of Arizona and elsewhere. Here, I will single out only two. Early on in my PhD, I made fast friends with fellow graduate students Christopher Howard and Santiago de Jesus Sanchez Borboa. Chris and Santi were both further along in the program than me, but they took me under their wings almost from day one, and we spent the better part of two years hanging out in Santi’s backyard, discussing every topic in philosophy under the hot, Tucson sun. These conversations were as crucial to my philosophical development as anything else I have done since coming to graduate school. And though we can, alas, no longer afford such an idyllic life style, Chris and Santi continue to provide me with unceasing personal and professional support to this day.

On this more personal note, I would also like to thank my parents for supporting me throughout my degree. I come from a family of lawyers, and I went off to undergrad assuming, like everyone else, that I would follow the same path. But when, in my final year of undergrad, I began seriously to consider the possibility of pursuing a PhD in philosophy instead, my father, Steven, immediately encouraged me to do so, telling me that he had once thought of doing the same but had chickened out, and hoped I would be braver. My mother,

Tammy, was initially less thrilled about the idea, especially since it meant I would have to move so far from home. But this only makes her unwavering support for me as I pursue this alternative path all the more meaningful.

Above all, however, I owe thanks to my fiancé Sarah Raskoff. I cannot begin to express all the ways she has helped and supported me throughout this process, not only as my best friend, but also as a colleague and astute philosophical interlocutor. So I will not even try. Instead, like much else, I dedicate this dissertation to her.

The bulk of Chapter 3 and a small part of Chapter 4 draw on my 2020 article, “Social Reform in a Complex World,” *Journal of Ethics and Social Philosophy* 17(2): 103-132. There are not enough high quality open-access journals in the discipline. Thanks to the editors at *JESP* for helping to blaze the way.

## TABLE OF CONTENTS

ABSTRACT.....	8
CHAPTER 1 – INTRODUCTION.....	10
1.1 Ideal Theory and its Critics.....	10
1.2 The Argument in Brief.....	14
1.3 Definitions and Clarifications.....	17
CHAPTER 2 – EVALUATING SOCIETIES.....	22
2.1 Comparative Evaluation and Social Reform.....	22
2.2 The Ideal Benchmark View.....	25
2.3 Sen’s Critique.....	30
2.4 The Redundancy Objection.....	38
2.5 Facilitation and Obstruction.....	53
CHAPTER 3 – CHASING TARGETS.....	61
3.1 Short-term versus Long-term Justice.....	61
3.2 Complexity and Problem Solving.....	64
3.3 The Ideal Target View.....	74
3.4 Complexity and the Ideal Target View.....	81
3.5 The Uses and Abuses of Ideal Theory.....	89
CHAPTER 4 – GETTING BETTER AT GETTING BETTER.....	95
4.1 A Methodological Challenge.....	95
4.2 From Problem Solving to Progressiveness.....	98
4.3 Progressiveness in Practice.....	107
4.4 Threading the Needle.....	112
4.5 Applying the Progressiveness Approach.....	128

CHAPTER 5 – CONCLUSION.....	145
5.1 The Mountain Analogy.....	145
5.2 Anti-Practicalism Revisited.....	150
5.3 Post-Script: Extensions.....	154
REFERENCES.....	156

## ABSTRACT

What, if anything, does our conception of the ideally just society tell us about how we should improve our own? I argue, against most ideal theorists, that it tells us very little: it provides us neither with a benchmark for evaluating nonideal societies nor with a long-term target for reform. But if the ideal can't guide us, then should political philosophers instead engage in problem solving—identifying present instances of injustice, diagnosing their causes, and prescribing targeted solutions—as some self-described nonideal theorists have recently suggested? While problem solving has its place, I argue that it is not enough: problem solving is a good way to identify short-term improvements in justice but not long-term ones. So how, then, should we theorize about long-term reform? I argue that we must supplement problem solving with an attempt to work out how to make our society more “progressive”: better at getting better, or more conducive to further improvements in general. I offer an account of progressiveness emphasizing experimentation, learning (epistemic feedback), and selection (practical feedback). And I begin to explore how we might improve the progressiveness of existing societies. Since, I suggest, one way to improve progressiveness is to shift the practice of political philosophy from a focus on the ideal society to a focus on problem solving and progressiveness, I conclude by arguing that we ought to make this shift not only so that we can better understand justice, but also so that we can better pursue it.

It is not the business of political philosophy and science to determine what the state in general should or must be. What they may do is aid in creation of methods such that experimentation may go on less blindly, less at the mercy of accident, more intelligently, so that men may learn from their errors and profit by their successes.

- John Dewey, *The Public and its Problems*

## CHAPTER 1 – INTRODUCTION

## 1.1 Ideal Theory and its Critics

Political philosophy has recently become self-conscious. Its dominant approach, ideal theory, is under siege. Leading the charge is a host of practical-minded critics, who allege that theorizing about the ideally just society is useless: that it tells us hardly anything about how to reform actual societies, about how to meet the injustices we currently face.<sup>1</sup> Some critics push further, arguing that ideal theory is not only useless, but harmful. Focusing too much on the ideally just society, they warn, blinds us to existing injustice; ideal theory might even serve as a pernicious ideology that reinforces an unjust status quo.<sup>2</sup> Others worry that ideal theory is liable not merely to stabilize injustice, but to inspire misguided reform efforts that only make things worse. Even the gradual pursuit of the ideal often comes at the expense of short-term justice. And if history is any guide, attempts to realize the ideal all at once through utopian revolution inevitably take a turn for the dystopian.<sup>3</sup>

Defenders of ideal theory respond to these critics in one of two ways. The first is to insist that ideal theory is worth doing even if it is useless or harmful: that it is a pernicious form of “practicalism” to assume that the value of intellectual work depends entirely on its practical relevance or practical effects.<sup>4</sup> Critics are right to highlight the limits of ideal theory, and to note the dangers of failing to recognize them. But within these limits, ideal theorists

---

<sup>1</sup> For recent arguments approaching this conclusion, see especially Anderson 2010, ch. 1, Gaus 2016, Mills 2005, Schmidtz 2011, Sen 2009, and Wiens 2015a. Historical antecedents include Dewey 1927, Hayek 2007, and Popper 2013.

<sup>2</sup> Mills 2005; see also Anderson 2010, ch. 1.

<sup>3</sup> Gaus 2016, Popper 2013.

<sup>4</sup> Estlund 2011b; see also Estlund 2011a, 2014.

are still engaged in a worthy endeavor; they are still attempting to understand something that is important to understand.<sup>5</sup> The rallying cry of these anti-practicalists might be: “the question of political philosophy is not what we should do but what we should think, even when what we should think makes no practical difference.”<sup>6</sup> They meet the charge that ideal theory is useless or harmful with a sidestep: “useless or harmful theory” does not imply “worthless theory,” so ideal theory might be worth doing all the same.

Other ideal theorists prefer to meet their critics head on, arguing that theorizing about the ideally just society is indeed useful because our conception of it provides us with a legitimate “aim to guide the course of social reform.”<sup>7</sup> On one specification of this view, ideal theory, so far from blinding us to injustice, helps us to identify ways of making things more just.<sup>8</sup> Our conception of the ideally just society may even serve as an indispensable *benchmark* of evaluation that we must outline before engaging in any serious theorizing about how to improve our current society.<sup>9</sup> On another, our conception of the ideal provides us with a *target* or long-term goal for reform.<sup>10</sup> Sure, some past attempts to pursue the ideal have been overhasty, and have exhibited a worrying tendency to ignore the short-term costs of doing so. But a few bad apples don’t spoil the bunch. We should avoid being reckless or

---

<sup>5</sup> Estlund 2011b.

<sup>6</sup> Cohen 2008, p. 268. Cohen does not defend ideal theory per se, but rather the value of theorizing about “fact-insensitive” normative principles that hold in all logically possible circumstances. But the sentiment is the same.

<sup>7</sup> Rawls 1999a, p. 15.

<sup>8</sup> Gilabert 2012, Stemplowska 2008, Swift 2008, Valentini 2011.

<sup>9</sup> Rawls 1999a; see also Shelby 2013.

<sup>10</sup> Rawls 1999b, Buchanan 2004, ch. 1, Robeyns 2008, Simmons 2010, Valentini 2012.

overzealous in pursuing the ideal, but we shouldn't ignore it either, lest our efforts to reform our society set back the achievement of the ideal, or worse, cut us off from it altogether.

These two lines of defense suggest that there are really two separate debates going on between ideal theorists and their critics. The first involves a proxy war over the more general question of the value of political philosophy and what role it should play in our society. On the one hand, we have the “practicalist” view that political philosophy should concern itself exclusively with practical matters—that an instance or type of political philosophy is valuable only if it helps make our society more just or contributes to a collective enterprise that serves that purpose. On the other, we have the “anti-practicalist” view that the practical relevance of political philosophy is but one consideration among others in determining its value—that one cannot indict a particular instance or type of political philosophy merely by showing that it is useless, because useless theorizing is often valuable. But since not even the most adamant anti-practicalist would insist that practical considerations are irrelevant to evaluating the *practice* of political philosophy—as it is carried out not only by academics, but by politicians, activists, educators, and the general public—the following should be common ground. If reorganizing or refocusing this practice would itself improve justice, then there are serious normative considerations in favor of making this change. If altering the way we do or think about political philosophy would make our society a more *progressive* one—one that more quickly or reliably improves in justice across time—then justice itself demands, or at least weighs heavily in favor of, such a change.<sup>11</sup>

---

Whereas the first debate concerns the extent to which political philosophy must be

<sup>11</sup> Compare Frazer 2016, who recasts the ideal/nonideal theory debate as a debate within the field of professional ethics over the obligations political philosophers have in virtue of “the moral demands of their vocation,” p. 178.

practically relevant, the second concerns the practical relevance of ideal theory. Here, there is the extreme view that ideal theory tells us absolutely nothing about how to improve our society—that it might even provide misleading guidance that stabilizes injustice or produces more of it. There is the other extreme view that we can't even begin to figure out how to reform our society without first completing an ideal theory—that ideal theory is a “necessary complement” to theorizing about social reform “without which the desire for change lacks an aim.”<sup>12</sup> And then there are more moderate views on which ideal theory is, in one way or another, relevant but not essential to theorizing about reform. When it comes to this second debate, however, there appears no neat way to bypass this disagreement by appealing to common ground. So we have no choice but to address the following two questions directly. First, does a conception of the ideally just society provide us with the guidance that ideal theorists claim: does it enable comparative evaluations of justice, provide us with a long-term goal for reform, or both? And, second, if it doesn't, then how should we theorize about, and pursue, social reform? If the ideal can't serve as our lodestar, then how are we to navigate?

My concern in this essay is primarily with this second debate: over the practical relevance of ideal theory. But the conclusions I reach on that topic will also help us to resolve the first debate—over whether and to what extent political philosophy should concern itself with practical matters. If we want political philosophy to be practically relevant, if we want it to play any significant role in guiding our efforts to promote reform, then, I will argue, political philosophy must undergo a major shift, from a focus on characterizing the ideally just society, to a focus on figuring out ways both of making our society more just, and of making it more *progressive*: better at getting better, or more conducive to further improvements in justice in general. And since, I will argue, one way to

---

<sup>12</sup> Rawls 2005, p. 285.

improve the progressiveness of our society is itself to modify political philosophy in precisely this way, we will find that there is a strong justice-based argument that we ought to make this change. We ought to change the way we do political philosophy not only so that we can better theorize about justice, but also so that we can better pursue it. Implementing this shift would itself amount to a genuine instance of social reform.

## 1.2 The Argument In Brief

I begin in Chapter 2 with the *ideal benchmark view*, on which ideal theory facilitates comparative evaluations of justice or the identification of ways of making our society more just.<sup>13</sup> This view, we will see, has a number of stronger and weaker variants, the strongest being that we cannot make any comparative evaluations of justice without an ideal in mind, and the weakest being that ideal theory can at least sometimes help us to figure out how to make our society more just. I reject strong versions of the ideal benchmark view outright, but acknowledge that ideal theory may nevertheless play some facilitating role in helping us to work out how to improve our society: it may aid us in developing our abstract criteria of comparative justice, may alert us to novel prospective reforms, and may help us to appreciate the ways that our society currently falls short. The usefulness of even these roles, however, is limited by the fact that an overreliance on ideal theory is liable to distort our criteria, to distract us from viable reforms, and to blind us to injustice, as critics rightly warn. So ideal theory should play only a fairly modest role in theorizing about how to improve our society's justice. It is not entirely useless, but it is not especially useful either.

---

<sup>13</sup> I adapt the distinction between the “ideal benchmark view” and what I will go on to call “the ideal target view” from Wiens 2015a, who draws a similar distinction between the “benchmark view” and the “target view.”

Given my rejection of the claim that we must identify the ideal before engaging in meaningful theorizing about how to improve our current society, an option presents itself. Perhaps theorizing about reform can instead take the form of *problem solving*—identifying extant problems of injustice, tracing out their causes, and proposing narrowly targeted solutions—as some critics of ideal theory have recently suggested.<sup>14</sup> But though problem solving has its use, ideal theorists argue that it is, by itself, unsatisfactory, because implementing solutions to the immediate problems we face may “retard,” “stall,” or “permanently block... movement toward overall justice.”<sup>15</sup> And this is why ideal theorists have been so attracted to another view about the practical role the ideal might play: that it serves as a long-term target whose pursuit we must trade off against the sort of short-term improvements to our society that we can discover through problem solving.

In Chapter 3, I turn to a discussion of both the *problem solving approach* and the *ideal target view*. Drawing on considerations of social complexity—the fact that societies are complex systems composed of many interacting parts—I show that ideal theorists are right to insist that even though implementing solutions to particular problems of injustice may be a good way to promote justice in the short term, this does not reliably promote overall justice in the long term. And I furthermore acknowledge with ideal theorists that if we could both confidently identify the ideally just society and confidently track our progress toward it, then ideal theory would provide us with a perfect supplement to problem solving. But I then argue that neither of these presumptions of ideal theory can be met—that the very complexity that renders problem solving unsatisfactory also undercuts our ability to identify the ideal or track our progress toward it, at least with sufficient confidence to warrant this

---

<sup>14</sup> See especially Anderson 2011, ch. 1, Schmidtz 2011, Sen 2009, Wiens 2012.

<sup>15</sup> Simmons 2010, p. 21.

pursuit. Furthermore, this complexity implies that there is most likely no fixed ideal to pursue in the first place, such that to chase the ideal is to chase a moving target. It would therefore be the height of folly to treat our current conception of the ideal as a long-term goal for reform to be pursued at the expense of short-term justice. If we wish to figure out how to pursue long-term justice, we need some alternative to ideal theory.

In Chapter 4, I develop this alternative. I argue that whereas ideal target theorists believe we must trade off short-term justice against *progress toward* the ideal, we should instead trade off short-term justice against the *progressiveness* of our society: the reliability and speed with which it continues to improve in justice indefinitely into the future. I offer a general account of progressiveness emphasizing a society's ability to flexibly implement many promising solutions to problems as they arise (experimentation), to stabilize those that prove successful while eliminating those that do not (selection), and to learn from both (learning). And I explicate and defend a *progressiveness approach* on which both problem solving and theorizing about progressiveness play important and complementary roles, but ideal theory plays hardly any role at all. Finally, I begin to explore how we might go about applying this approach, and argue, among other things, that changing the way we practice political philosophy would itself contribute to the creation of a more progressive society, and would therefore contribute, in the long run, to a more just society as well.

I conclude in Chapter 5 by arguing for a shift in political philosophy from the dominance of ideal theory to the widespread adoption of the progressiveness approach. Here, I provide a final statement of my response to the anti-practicalist defense of ideal theory, hinted at above, that this shift—which significantly deemphasizes but does not altogether do away with ideal theory—is required both so that we can better understand justice, and so that we can better pursue it. I end with a brief post-script on the extent to

which the conclusions of this essay might extend to other domains, forcing us to reconsider how we think about social morality, personal virtue, and rationality. In each case, I conjecture that we would be better served by focusing less on ideals, and more on problem solving and progressiveness.

### 1.3 Definitions and Clarifications

To this point, I have been using terms like “ideal” without defining them, allowing context to do the work for me. Carried on too long, this is a recipe for confusion, since many come to the ideal/nonideal theory debate with different preconceptions about the meaning of its central terms. So in an effort to avoid misunderstanding, I will here provide some brief definitions and clarifications of those terms that would otherwise be most likely to mislead.

Though there are many ways to distinguish between “ideal” and “nonideal” theory, the distinction I have in mind throughout this essay is the one originally introduced by John Rawls, between ideal theorizing about the “ideally” or “perfectly just society” and nonideal theorizing about “how we are to deal with injustice.”<sup>16</sup> Other ways of drawing the distinction—for example, between full compliance theory (in which we assume that everyone complies with demands of justice) and partial compliance theory (in which we don’t), or theory that idealizes (a lot, or in certain ways) and theory that doesn’t (very much, or in those particular ways)—are not my concern, except to the extent that they bear on this first distinction.<sup>17</sup> So each time I use the term “ideal theory,” I mean “theory that aims to

---

<sup>16</sup> Rawls 1999a, p. 8.

<sup>17</sup> For example, if, as Rawls suggests, everyone in the ideally just society would comply with the demands of justice, then the ideally just society would be a full compliance society. But, as I will explain, there are other conceptions of the ideally just society on which it need not

characterize an ideally just society,” and each time I use the term “nonideal theory,” I mean “theorizing about how we are to deal with injustice.” In general, however, I avoid the label “nonideal theory,” instead preferring to call theorizing about how to reform our current society “reform theory.” One reason for this is that theorizing about how to deal with injustice involves a broader range of topics than theorizing about how to reform our society—including questions of how we are to compensate victims of injustice and what obligations we, as individuals, have to promote social reform—that are not my concern here.<sup>18</sup> Another is that the common practice of referring to theorizing about how to reform our society as nonideal theory—as theorizing that is *not* about the ideal—only reinforces the view that ideal theory is the dominant approach in political philosophy, departures from which define other approaches. And this is precisely what I seek to overturn.

Now, in the interest of style, I sometimes refer to the ideally just society as the “ideal society,” or simply as the “ideal.” It is worth mentioning this here to forestall two possible misunderstandings. First, one might think that there is a difference between the ideally just society and the ideal society, since there are other values besides justice that are relevant to our overall evaluation of a society.<sup>19</sup> But, be that as it may, I will generally pass over such distinctions here, and speak instead as if justice depends on all the morally or politically relevant considerations that plausibly bear on our evaluation of a society, including equality, freedom, desert, need, efficiency, rights, oppression, discrimination, exploitation, 

---

involve full compliance. For further discussion of the various ways that Rawls and others draw the ideal/nonideal contrast, see Hamlin and Stemplowska 2012 and Valentini 2012.

<sup>18</sup> Compare Shelby 2016, p. 12.

<sup>19</sup> Cohen 2008 is most famous for emphasizing this point, but it was already present in Rawls 1999a, p. 9.

domination, and so on. Some might object to this all-encompassing understanding of justice, arguing, for example, that we must distinguish between the justice of a society (which depends, say, on questions of distribution) and the extent to which it fulfills humanitarian obligations (which depends, say, on questions of need). But while such distinctions are important in other contexts, they are distractions here, since none of my arguments depend on how narrowly or widely we draw the contours of our conception of justice, and proceeding as if we draw them quite widely vastly simplifies our discussion. Furthermore, while I focus on reforming particular societies rather than the international order, I do not assume that questions of international justice are irrelevant to evaluating the justice of societies. For example, I allow that the justice of a society might depend on its effects on other societies through its policies concerning immigration, trade, or foreign aid.

Second, values such as equality, freedom, and efficiency are sometimes themselves referred to as “ideals.” I have no objection to this use of language, but—to avoid ambiguity—I will never employ it myself, and will instead always refer to such values as “values,” reserving the term “ideal” as shorthand for “the ideally just society.”<sup>20</sup> The ideal, then, is whichever society best or most fully satisfies the values that are relevant to justice: it is the society that is most just. So when I discuss ideal theory, I do not have theorizing about abstract values such as freedom and equality in mind, nor do I mean theorizing about abstract criteria of comparative justice that allow us to rank societies from more to less just. Instead, I am referring to theorizing about how to arrange society so that such values are best satisfied, or it ranks highest according to such criteria. As we will see, such ideal theorizing may involve an attempt to specify either the abstract principles that would be

---

<sup>20</sup> Compare Hamlin and Stemplowska 2012 on the distinction between “ideal theory” and a “theory of ideals.”

realized in the ideal society or the institutional arrangements and other concrete features the ideal would contain. Indeed, the ideal society, like any society, may be described at various levels of abstraction. And different levels may be appropriate in different contexts.

The careful reader may have noticed, however, a remaining ambiguity in the notion of the ideally just society. I have said that the ideally just society is the most just society, but this claim makes implicit reference to a comparison class. By the “ideally just society,” do I mean the most just society that is logically possible, the most just society that is possible given the laws of physics and human psychology, the most just society that we could ever achieve from where we are, or what? Though it is important to keep this question in mind, for now, I leave it deliberately unsettled. One of the lessons of this essay will be that the role the ideal might play depends essentially on how we specify the relevant comparison class. For example, we will see that the view that we ought to treat the ideal as a long-term goal is only remotely plausible if we have something like “the most just social arrangement that we could achieve from where we are” in mind, whereas the view that the ideal serves as a benchmark of evaluation is subject to no similar constraint. And this, by the way, is why I generally employ the term “ideally just society” rather than “perfectly just society” or “fully just society.” Depending on how we fill out the relevant comparison class, the ideally just society might fall short of the standard of conceptual perfection that is suggested by the label “perfectly” or “fully” just. For example, if by “ideally just society” we mean “most just society compatible with certain facts about human nature,” and if one such fact is that there will always be some subset of the population that will act unjustly, then even the ideally just society may contain instances of injustice perpetrated by such incorrigibles.

These clarifications about how I use the term “ideal” should suffice for now. But there remains one other term that warrants discussion. This is the word “progressive” as

appears in the subtitle of this essay, and as I have already employed a few times. As I use the term, the progressiveness of a society depends on how quickly and reliably it improves in justice across time. It depends on the *velocity* with which a society traverses the space of justice (its direction and rate of change) rather than its current *position* in that space (how just it is). So, for example, if the justice of a society depends on its wealth, then a society's progressiveness depends on its rate of economic growth. If justice tracks economic equality, then progressiveness tracks the rate at which a society decreases in inequality across time. And if justice depends on the extent to which a society satisfies a variety of economic and non-economic values, then a society's progressiveness depends on the rate at which it comes to better satisfy these values over time, rather than on their current degree of satisfaction.

It is an unfortunate fact of the English language that it contains no word that captures precisely this concept, and that the closest and most versatile word available—"progressive"—has come to be associated, at least in many places, with a particular left-leaning political orientation. So let me be as clear as I possibly can about this. When I say "progressive" I do not have this or any other particular political orientation in mind, nor will I defend any such orientation in this essay. This work offers no substantive theory of justice, no substantive theory of which fundamental values or criteria we ought to employ when evaluating current or hypothetical societies. What it offers is a general theory of how we ought to theorize about and pursue justice, which, I argue, one ought to accept regardless of what fundamental values or criteria one brings to the table. It should therefore be of interest to anyone interested in justice, regardless of whether one endorses liberal egalitarianism, luck egalitarianism, libertarianism, utilitarianism, sufficientarianism, perfectionism, republicanism, any other –ism, or no particular –ism at all.

## CHAPTER 2 – EVALUATING SOCIETIES

### 2.1. Comparative Evaluations and Social Reform

Ideal theorists aim to describe the ideally just society. They make judgments about the principles that would be realized in that society, or about the institutions and other social features it would contain. On their face, such judgments provide us with little guidance. Except in the mythical case where we can realize the ideal in a single stroke, judgments about the most just society seem irrelevant to which changes we should implement here and now. To figure out how to reform our current society, we must instead rely on comparative evaluations of justice. We must ask which changes to our society would produce a *more just* society, not what the *most just* society would be like.

On the basis of this sort of reasoning, Amartya Sen has recently argued that ideal theory is irrelevant to theorizing about social reform.<sup>21</sup> Although Sen does not put things in precisely these terms, we may reconstruct his basic argument as proceeding in two steps. First, Sen defends what I will call a *problem solving approach* to theorizing about reform, on which the sole goal of reform theory is to identify concrete changes to our current society that would make it more just. The goal, in other words, is to identify *problems* or “remediable injustices” in our current society—features that render it less just than it would otherwise be—as well as *solutions* to these problems, or concrete changes that would eliminate or at least mitigate problems of injustice, thereby making our society more just.<sup>22</sup> Second, Sen argues that our conception of the ideally just society does not help us to engage in such problem solving: it is not necessary, sufficient, or even useful for this purpose. So ideal theory doesn’t tell us anything about which changes to our society would make it more just,

---

<sup>21</sup> Sen 2009, esp. ch. 4.

<sup>22</sup> Ibid., p. viii, *et passim*.

and, if we endorse the problem solving approach, it is therefore irrelevant to reform theory. At best, it amounts to “an engaging intellectual exercise in itself.”<sup>23</sup>

Sen’s argument is controversial, but our reconstruction of it provides us with a fruitful starting point because each of its steps highlights a different ground on which ideal theorists resist Sen’s conclusion and affirm the practical relevance of ideal theory. Some challenge the first step, calling into question the adequacy of the problem solving approach. To responsibly theorize about how to reform our society, they argue, we cannot only solve particular problems of injustice, but must also take into account the extent to which “solutions” to these problems set back or further our long-term goal or *target* of eventually achieving the ideal; and for this, we need ideal theory.<sup>24</sup> Others object to the second step, arguing that problem solving requires, or at least is greatly facilitated by ideal theory. A conception of the ideally just society may even serve as an indispensable *benchmark* for evaluating societies in terms of how far they fall from it. It may provide, as Rawls famously puts it, “the only basis for the systematic grasp of these more pressing problems.”<sup>25</sup>

In a sense, it is the first challenge to Sen’s argument that is more fundamental. If problem solving is a flawed approach for theorizing about reform, then even if Sen and other critics are right that ideal theory doesn’t help much with problem solving—well, so what? Ideal theory may remain an essential component of a more sophisticated approach to reform theory that attempts to work out how to realize the ideal as a long-term target. But even on this *ideal target view* problem solving retains an important role. In evaluating prospective changes to our society, we must consider not only the extent to which they

---

<sup>23</sup> Ibid., p. 101.

<sup>24</sup> Rawls 1999b, Buchanan 2004, ch. 1, Robeyns 2008, Simmons 2010, Valentini 2012.

<sup>25</sup> Rawls 1999a, p. 8.

would promote or set back the eventual achievement of the ideal, but must trade this off against the extent to which they promote or set back short-term justice by mitigating or aggravating the problems we currently face. And, as we will see, every plausible approach to reform theory similarly leaves room for problem solving in this way. So the question is not whether problem solving should be one component of reform theory, but whether it should be the only component. And, seen in this light, it is the second challenge to Sen's argument that is more urgent. If problem solving itself relied on ideal theory, then no matter what we think about the adequacy of the problem solving approach, we would still need to engage in ideal theory. Any plausible approach to reform theory would require it.

In this chapter, I therefore postpone questions about the adequacy of the problem solving approach, confining my attention to the question of whether ideal theory is in some sense prior to problem solving or otherwise facilitates it. I begin with an explanation of the *ideal benchmark view* and of Rawls's claim that it implies a certain priority of ideal theory over problem solving, along the way showing why a common criticism of the view—namely, that we are often able to identify improvements in justice without any ideal in mind—rests on a misunderstanding. I then turn to Sen's own criticisms of the ideal benchmark view, which I also find to be lacking, but which I draw on to develop a more powerful *redundancy objection* to the benchmark view and the claim of priority it allegedly entails. My argument, in brief, is that the ideal benchmark view commits one to thinking that the truth of particular judgments of “more just” or “less just” are explained by *ideal deviation* criteria of justice on which one society is less just than another when it deviates further from the ideal along some given dimension. But it turns out that all plausible ideal deviation criteria make the ideal redundant, in the following sense: we can use the dimensions picked out by such criteria to make comparative evaluations of justice directly, without ever referring to the ideal.

With my argument against the ideal benchmark view complete, I turn to an examination of the extent to which ideal theory might facilitate problem solving even without being prior to it. The conclusion I reach is that while ideal theory can indeed play some subsidiary role in problem solving—it may help us to develop abstract criteria of comparative justice, to come up with creative solutions to the problems we face, and to better appreciate such problems—it is not central to the approach. The stage will then be set for a more careful consideration of the problem solving approach, ideal target theorists’ rejection of it on the grounds that we must also attempt to work out how to make progress toward the ideal, and my own alternative approach to reform theory that acknowledges the needs to supplement problem solving, not with ideal theory, but with an attempt to work out how to make our society more progressive.

## 2.2. The Ideal Benchmark View

Like much else in contemporary political philosophy, recent interest in the practical relevance of ideal theory traces to John Rawls. Rawls’s own views on the matter are complicated, and arguably changed as his career progressed. In later work, he explicitly endorsed the ideal target view.<sup>26</sup> But in *A Theory of Justice*, Rawls proclaimed that our conception of the ideally just society—and, in particular, of the principles that would govern its basic structure or major social and political institutions—is practically relevant because “[e]xisting institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason.”<sup>27</sup> Tommie Shelby explains:

---

<sup>26</sup> See the various quotes in Chapter 3 below.

<sup>27</sup> Rawls 1999a, p. 216.

On the Rawlsian view, injustices are conceptualized as *deviations* from the ideal principles of justice, in much the same way that fallacious reasoning is conceived as a deviation from the rules of logical inference. An injustice is a failure on the part of individuals or social arrangements to satisfy what the ideal principles of justice demand. Thus, charges of injustice *presuppose* ideals of justice, which particular individuals and institutions can and often do depart from. Such deviations can be small or great, minor or serious... depending on the size and nature of the gap between ideals and practice.<sup>28</sup>

So, on this ideal benchmark view, injustice is a deviation from ideal justice, and for one society to be less just than another is for it to deviate further from the ideally just society. Much as something is a worse replica of an original painting to the extent that it deviates from that painting, a society is less just to the extent that it deviates from the ideal society.

Critics of ideal theory often assume that the ideal benchmark view implies a strong claim of epistemic priority: that “we first need to know what an ideally just society would be, to identify the ways our current society falls short,” and, therefore, to identify ways of making it more just.<sup>29</sup> Their standard move is then to point to various obvious instances of

---

<sup>28</sup> Shelby 2013, p. 153.

<sup>29</sup> Anderson 2009, p. 3. Though Sen is often held to interpret the benchmark view in this way—and though many of his remarks lend themselves to such a reading—I do not include him in this group because, in his careful moments, he clarifies that he is concerned with a more sophisticated version of the ideal benchmark view, on which “the transcendental question (‘what is a[n ideally] just society?’) has to be answered first, as an essential requirement, for a cogent and well-founded *theory* of comparative justice” (Sen 2009, p. 98, emphasis mine). I return to this contrast between epistemic and theoretical priority shortly.

injustice whose elimination would just as obviously improve justice—for example, slavery, caste and gender discrimination, and enforced racial segregation—to note that we can judge as much without referring to the ideal, and to declare the ideal benchmark view defeated. Such examples show, after all, that we can identify problems of injustice without having an ideal in mind, and that we “can judge social option A as being better than social option B without starting with a view of the best society.”<sup>30</sup> Indeed, if we know anything at all about justice, it is that slavery is unjust and that abolishing it makes things more just. At least when it comes to justice, then, “knowledge of the better does not require knowledge of the best.”<sup>31</sup>

These critics are right. We typically identify injustice long before we have any conception of ideal justice; we “recognize the existence of a problem before we have any idea of what would be best or most just.”<sup>32</sup> But the very obviousness of this claim should give us pause—could Rawls and other ideal benchmark theorists really think otherwise? As far as I am aware, the answer is “no”: no actual ideal benchmark theorist holds that we cannot identify any problems of injustice, that we cannot make any comparative evaluations of justice, without an ideal in mind. Rawls himself, for example, held that theorizing about justice involves a process of moving “back and forth” between (among other things) our judgments about particular instances of injustice and our judgments about the principles that would govern the ideally just society—sometimes revising the former in light of the latter, and sometimes revising the latter in light of the former, until all our judgments hang together in “reflective equilibrium.”<sup>33</sup> Our judgment that slavery is very unjust, for example, serves as

---

<sup>30</sup> Appiah, 2017, p. 168.

<sup>31</sup> Anderson 2009, p. 3.

<sup>32</sup> *Ibid.*, p. 3.

<sup>33</sup> Rawls 1999a, p. 18, *et passim*.

a “fixed point” in our theorizing about justice; any plausible conception of the ideally just society must accommodate the datum that slavery is (very) unjust and therefore deviates (greatly) from the ideal.<sup>34</sup> So Rawls himself should not be saddled with the view that we must form a conception of the ideal before we can make any judgments about what would improve our society’s justice. And neither should the ideal benchmark view more generally.

Instead, the ideal benchmark view claims that our conception of the ideally just society provides a benchmark of evaluation “comparison with which defines a *standard* for judging actual institutions.”<sup>35</sup> In other words, the view is that our *criteria* of comparative justice make essential reference to the ideal, since such criteria must be formulated in the following *ideal deviation* form: one society is less just than another if and only if it deviates further from the ideal.<sup>36</sup> And from here, it is not hard to see why ideal benchmark theorists insist that ideal theory enjoys a certain theoretical priority over problem solving, in the sense that the latter must be “worked out after an ideal conception of justice has been chosen.”<sup>37</sup> The goal of problem solving is, after all, to identify concrete ways of making our society more just. And though, as we have seen, there may be some easy cases where we can tell that a change would improve justice without needing to resort to any theory, problem solving is presumably supposed to help us get beyond these cases. But to do this, it is commonly thought that we must invoke criteria of comparative justice—criteria of the form “society *x* is more just than society *y* if and only if...” —that explain why particular changes to our

---

<sup>34</sup> Rawls 2001, p. 29.

<sup>35</sup> Rawls 1999a, p. 199, emphasis mine.

<sup>36</sup> Compare Sen 2009, p. 98, who employs the term “distance-comparison” instead of “ideal deviation.”

<sup>37</sup> Rawls 1999a, p. 8.

society would make it more (or less) just. So if ideal benchmark theorists are right that criteria of comparative justice make essential reference to the ideal, and if problem solving requires us to appeal to such criteria, then it follows that we cannot engage in problem solving without referring to the ideal. Though we may make scattered pretheoretical judgments about which changes to our society would improve things without any conception of the ideal in mind, we cannot engage in serious theorizing about how to improve our society until we have first completed our ideal theory.<sup>38</sup>

One way to respond to this claim of theoretical priority would therefore be to deny that we need criteria of comparative justice to engage in problem solving in the first place. For example, one might defend a version of methodological particularism on which, regardless of whether the ideal benchmark view is right that criteria of comparative justice make essential reference to the ideal, such criteria are unnecessary or unhelpful when theorizing about how to improve our society: we can get by simply by relying on our pretheoretical convictions about justice.<sup>39</sup> But while I am sympathetic to this response, at least to a point, I will set it aside for now, and grant that we need criteria of comparative justice to engage in problem solving. In what follows, then, I will tackle head on the question of whether ideal benchmark theorists are right that criteria of comparative justice make essential reference to the ideal, assuming that, if they are, the claim of theoretical priority follows. Sen's own criticisms of the ideal benchmark view, though ultimately unsuccessful, provide a useful starting point.

---

<sup>38</sup> Compare Valentini 2011, Shelby 2013, p. 156.

<sup>39</sup> I adapt the concept of "methodological particularism" from Sinnott-Armstrong 1999, who helpfully contrasts it with various other forms of particularism.

### 2.3. Sen's Critique

According to the ideal benchmark view, criteria of comparative justice make essential reference to the ideal. This makes such criteria fairly atypical, because, as Sen has remarked, most criteria of comparative evaluation do not work this way. To take Sen's most famous example, we don't need to refer to the highest mountain to make a principled comparison of the heights of Mount Kilimanjaro and Mount McKinley.<sup>40</sup> We can instead rely, say, on the criterion of which mountain is longer from its base to its summit. And, similarly, we don't need to refer to the most beautiful painting to make a principled comparison of the beauty of a Picasso and a Dali.<sup>41</sup> Though a theory of comparative beauty is no doubt more complicated than a theory of height, it is likely to take into account some number of criteria of comparative beauty—concerning, for example, symmetry, form, and proportion—to assign rules for balancing these criteria in cases where they conflict, and to claim that the more beautiful of two paintings is the one that better satisfies these criteria given these balancing rules. Alternatively, it may simply say that one painting is more beautiful than another when it tends to produce more aesthetic appreciation in its viewers than another, at least under certain viewing conditions. Of course, other options are available too.

Sen takes such examples to show that ideals play no necessary role in comparative evaluation. If criteria of comparative height make no reference to the highest point, and criteria of comparative beauty make no reference to the most beautiful object, then why should criteria of comparative justice make any reference to the most just society? But this argument is, by itself, too quick. Though Sen's examples do show that "[t]here would be something deeply odd in a *general* belief that a comparison of any two alternatives cannot be

---

<sup>40</sup> Sen 2009, p. 102.

<sup>41</sup> Ibid, pp. 16, 101.

sensibly made without a prior identification of a supreme alternative,” ideal benchmark theorists are not committed to any such general belief.<sup>42</sup> Instead, they need claim only that *some* comparative criteria make essential reference to a supreme alternative, and that the correct criteria of comparative justice belong in this class. If the ideal benchmark view holds, then criteria of comparative justice are more like criteria of how far a point deviates from a particular mountain, or how far a replica deviates from an original painting, than criteria of how high or beautiful an object is. The correct criteria of comparative justice must be formulated in ideal deviation form, and therefore make essential reference to the ideal, even if other candidate criteria of comparative justice need not.

The real upshot of Sen’s argument, then, is that it is a substantive question whether criteria of comparative justice make essential reference to the ideal—not one that can be settled by reflecting on the logic of comparative evaluation in general. So we have two positions in front of us. On the first, which Sen endorses, criteria of comparative justice are *essentially comparative*: they can yield verdicts about which of two societies is more just than the other simply by referring to the features of those two societies. This is compatible with a monistic theory on which only a single essentially comparative criterion is relevant to justice—say, a utilitarian view on which one society is more just than another when it has greater well-being in it than the other, or a perfectionist view on which one society is more just than another if it exhibits more human excellence. And it is also compatible with a pluralistic theory, on which one society is more just than another when it better satisfies a plurality of essentially comparative criteria—concerning, say, which society is more efficient, equal, or free than the other—given rules for balancing these criteria in cases where they conflict. On the second, which ideal benchmark theorists endorse, comparative evaluations

---

<sup>42</sup> Ibid., p. 102, emphasis mine.

of justice are instead explained by one or more *ideal deviation* criteria that make essential reference to the ideal, in the following sense. They can yield verdicts about which of two societies is more just than the other only by also referring to the ideally just society, and asking which departs further from this ideal.<sup>43</sup>

With this in mind, let us turn to Sen's second worry, namely, that such ideal deviation criteria are not very plausible, because "descriptive closeness is not necessarily a guide to valuational proximity": for example, "a person who prefers red wine to white may prefer either to a mixture of the two, even though the mixture is, in an obvious descriptive sense, closer to the preferred red wine than pure white wine would be."<sup>44</sup> In other words, Sen's worry is the familiar one that, of three options, the option that deviates less from the best option is not necessarily "second-best" of the three.<sup>45</sup> To see this in the case of justice, suppose we understand deviations in terms of institutional similarity: a society deviates from

---

<sup>43</sup> The possibility of such ideal deviation criteria is overlooked in Wiens' 2015a otherwise persuasive criticism of the ideal benchmark view. Wiens argues that the ideal is redundant when making comparative evaluations of justice, but his argument takes as a premise a "general model of normative political theory" on which basic criteria of comparative justice are all essentially comparative, pp. 427-428. Later, I will attempt to vindicate both Wiens's premise and his conclusion, but without simply assuming the premise from the outset.

<sup>44</sup> Sen 2009, p. 16.

<sup>45</sup> See Lipsey and Lancaster 1956, whose "general theory of second best" I discuss in more detail in the next chapter. Although Lipsey and Lancaster were themselves concerned with showing that deviating less from an ideally efficient market does not necessarily improve efficiency, their basic idea generalizes. See, for example, Goodin 1995 for a helpful discussion in the context of political philosophy.

the ideal society to the extent that its institutions differ from those that would be realized in the ideal. Now suppose further that the ideal would contain no oppression, and so no institutions in place to mitigate oppression or compensate its victims. Abolishing the ameliorative or compensatory institutions that we have in place would therefore make our society more institutionally similar to the ideal—but surely this would not make things more just!<sup>46</sup> Or, to take a case more analogous to Sen’s wine example, consider healthcare. Though the position rests on contentious empirical claims, there is nothing inconsistent in holding that the ideal society would contain a purely market-based healthcare system—say, because it is most efficient—or a purely public system—say, because it is most equal—while also insisting that both pure options are more just than a “mixed” or two-tier system—say, because it is worse with respect to both efficiency and equality. So, again, deviating further from ideal institutions does not necessarily make a society less just.

Now, an obvious response to this second worry is that I have simply employed the wrong deviation measure. There are all sorts of ways to measure deviations from the ideal, and these examples merely show that institutional deviations are not what is relevant. Indeed, as Pablo Gilabert points out, something similar is true of Sen’s wine example:

[P]erhaps a mixture of red and white wine will be descriptively closer to pure red wine, in one respect, than pure white wine would be. But this descriptive similarity may not be important, or decisive, for the agent making the comparisons, because the important principles rendering red wine best would be better instantiated in white wine than in the mixture. Perhaps color is not evaluatively important, while consistency of taste is.<sup>47</sup>

---

<sup>46</sup> Estlund 2014, pp. 120-121.

<sup>47</sup> Gilabert 2012, p. 46.

In other words, Gilabert's suggestion is that we can avoid second-best worries in the case of wine by ensuring that we measure deviations with respect to those features of my favorite wine that most fundamentally explain why I like it best, rather than its more incidental features. If, for example, I like a particular red wine best exclusively because of its taste, and care not at all about its color, then if a deviation measure is to track how much I like wine, it should measure departures with respect to taste, not color. And returning to the case of justice, Gilabert argues that we can avoid second-best worries in much the same way. Though features like taste are clearly irrelevant to justice, there are certain basic values or principles whose implementation by ideal institutions is what explains why those institutions would be contained in the ideally just society.<sup>48</sup> Nonideal societies are those that fall short of realizing these basic values or principles, and an adequate ideal deviation criterion should judge that a society is less just to the extent that it falls further short in this regard. So, to return to our earlier example, if what explains why the ideal would contain a public healthcare system (if it would) is the fact that this system would realize a principle mandating the equal distribution of healthcare resources (if this is so), then, everything else being equal, a society that comes closer to realizing this principle of equal distribution should count as more just than one that doesn't—even if it does so by deviating further from the ideal society's institutions.

Gilabert provides a compelling response to Sen's second challenge. As Sen himself stresses, the justice of institutions is explained (at least to a significant extent) by their "social realizations": by their effects on further features of society, for example, its distribution of resources, that are fundamentally relevant to that society's justice.<sup>49</sup> So, in order to pick out a

---

<sup>48</sup> Gilabert 2012, p. 48.

<sup>49</sup> Sen 2009, esp. ch. 10.

particular institutional arrangement as ideal, we must invoke some basic values or principles that explain what it takes for an institutional arrangement to be ideal—for example, the fact that it produces an equal distribution of resources. And because institutions *interact* with one another to produce their justice-relevant effects (a topic we will return to at much greater length in Chapter 3), a society that is closer to the ideal with respect to its institutions may fall further from it with respect to the realization of these basic values or principles—for example, by distributing resources less equally. This is what generates second-best problems when we attempt to measure deviations with respect to institutional similarity, but there is no room for such problems to arise if we measure deviations with respect to basic values or principles themselves. We can thus avoid second-best problems, and therefore Sen’s second worry, by ensuring that our ideal deviation criteria track deviations in this latter way. They must count one society as deviating further from the ideal when it falls further short of satisfying those basic values or principles, whatever they are, whose optimal satisfaction by an institutional arrangement is what fundamentally explains why that arrangement is ideal.<sup>50</sup>

This brings us to Sen’s third and final objection, which calls into question the sufficiency rather than the necessity or plausibility of appealing to the ideal when employing criteria of comparative justice:

The difficulty lies in the fact that there are different features involved in identifying distance [from the ideal], related, among other distinctions, to different fields of departure, varying dimensionalities of transgressions, and diverse ways of weighing separate infractions. The identification of [the ideal] does not yield any means of addressing these problems to arrive at a relational ranking of departures... The characterization of spotless justice, even if such a characterization were to emerge

---

<sup>50</sup> Compare Swift 2008, pp. 372-378.

clearly, would not entail any delineation whatever of how diverse departures from spotlessness would be compared and ranked.<sup>51</sup>

As an illustration, Sen points out that “in a Rawlsian analysis... departures may occur in many different areas, including the breaching of liberty, which, furthermore, can involve diverse violations of distinctive liberties”; that “[t]here can also be violations—again, possibly in disparate forms—of the demands of equity in the distribution of primary goods”; and that “[w]e have to consider, further, departures in procedural equality.”<sup>52</sup> Or, to take another example, suppose we identify the ideal as a society in which no one is oppressed. One obvious way for a society to deviate from this ideal is for it to contain more total oppression. But simply saying that the ideal contains no oppression does not tell us how to measure or aggregate different sorts of oppression to determine which of two societies deviates further from the ideal in this regard. Further, if no one is oppressed, then oppression is equal, so another way for a society to deviate from this ideal is for it to have a less equal distribution of oppression. But is this relevant, and, if so, how are we to measure such deviations and balance them against deviations with respect to total oppression? And these are only two of the many sorts of deviations we might consider. For example, perhaps we should instead measure deviations with respect to the level of oppression experienced by the most oppressed member of society, who would experience no oppression in the ideal.

Sen’s worry here, to be clear, is not that it is impossible to construct deviation measures that permit us to provide overall evaluations of how far a society deviates from the ideal: we can, after all, stipulate whatever deviation measure we like. It is rather that we can

---

<sup>51</sup> Sen 2009, p. 98-99.

<sup>52</sup> *Ibid.*, p. 99.

opt for many different deviation measures, and simply specifying an ideal does not tell us which to pick. With only a conception of the ideal in mind, we can judge that all nonideal societies are less just than the ideal, but we cannot make any further judgments about the relative justice of nonideal societies. For that, we must specify further which departures from the ideal are relevant to justice, and in which dimensions we should measure these deviations. If only one dimension of deviation is relevant, this will provide us with a single ideal deviation criterion that claims that one society is more just than another when it deviates less from the ideal in this dimension. But if more than one dimension is relevant, we will then have a set of ideal deviation criteria, and will need to provide further rules for balancing these different sorts of departures against one another in cases where they conflict.

Once again, Sen's criticism yields an important lesson: that a conception of the ideal cannot, by itself, yield comparative evaluations of nonideal options. But his conclusion that we must reject the ideal benchmark view does not follow. Though ideal benchmark theorists claim that criteria of comparative justice make essential reference to the ideal, they need not claim further that specifying a conception of the ideal is all we need to generate a ranking of nonideal options. Instead, they can acknowledge that ideal deviation criteria must include two elements: a conception of the ideally just society, and a deviation measure that specifies some dimension along which deviations from the ideal are fundamentally relevant to justice. Without employing such a deviation measure we cannot determine the extent to which a society deviates from ideal justice, and so cannot make comparative evaluations of nonideal alternatives. But all this remains compatible with ideal benchmark theorists' insistence that such criteria also require us to specify some conception of the ideal from which we must measure how far societies deviate in the relevant dimension, and that such criteria therefore make essential reference to the ideal.

#### 2.4. The Redundancy Objection

So much for Sen's critique. There remains, however, a deeper worry in the vicinity, which I will now argue ultimately derails the ideal benchmark view. The worry is that once we fully specify a plausible ideal deviation criterion, we make the ideal redundant: we end up with a criterion that is equivalent to an essentially comparative one, and that therefore makes no essential reference to the ideal at all. For notice where our replies to Sen have left us. If the ideal benchmark view holds, then comparative evaluations of justice are explained by ideal deviation criteria (along, perhaps, with rules for balancing them) that each contain two components: a conception of the ideal, and a deviation measure that specifies some fundamentally justice-relevant dimension along which societies can come closer to or further from realizing this ideal. But if this is so, then once we have a plausible deviation measure in place, it appears to do all the work. We can use the dimension picked out by our deviation measure to make comparative evaluations directly, without ever needing to make any reference to the ideal. Or so I will argue—first, by applying this redundancy objection to a number of views in the literature, and then by showing why it generalizes.

##### *Applying the Redundancy Objection*

To begin to get a feel for the redundancy objection, consider a toy example. Suppose you defend an institutional arrangement—say, a particular specification of the classical liberal state—as ideal, and, since you are an ideal benchmark theorist, you are convinced that criteria of comparative justice must be formulated in ideal deviation form. But you are then faced with the question of how to measure deviations from the ideal, so you turn to the prior question of what basic values or principles explain why this institutional arrangement is ideal. After careful consideration, you come to the following conclusion: the classical liberal state is

ideal because it promotes more total well-being than all other feasible arrangements. Further, you decide that since, on your view, total well-being is all that justice is fundamentally concerned with, you should measure deviations from the ideal in the dimension that arranges societies from less to more total well-being. And so you come to the following ideal deviation criterion: a feasible society is more just than another when it differs less in its total well-being from the ideally just society. But you now realize that since the ideally just society has the most well-being of any feasible society, your principle need not be formulated in ideal deviation form after all, because it is equivalent to the essentially comparative utilitarian criterion that one feasible society is more just than another when it has a greater sum of well-being in it. Though you began as an ideal benchmark theorist, you ended up with a criterion that makes no essential reference to the ideal. It makes the ideal redundant.

I call this a toy example because no one would really expect a utilitarian to accept the ideal benchmark view. Utilitarianism is a quintessential essentially comparative theory of justice, because it claims that one society is more just than another when it produces more total-well-being than the other, and so simply picks out the ideal as whichever society is more just than all other feasible societies given this prior essentially comparative criterion. But, as we will now see, many actual ideal benchmark theorists fall prey to precisely the same objection as our hypothetical utilitarian. They begin with a conception of the ideally just society, specify some dimension  $d$  in which we can arrange societies from less to more  $d$ , and then defend an ideal deviation criterion on which one society is less just to the extent that it deviates further from the ideal in dimension  $d$ . But this criterion then turns out to be equivalent to the essentially comparative criterion that a society is more just whenever it is more (or whenever it is less)  $d$ , thus rendering the ideal redundant.

Consider, for example, the most explicit attempt to spell out an ideal deviation

criterion in the literature: Thomas Christiano and Will Braynen’s “common good” principle of equality, on which the justice of a society depends on its deviation from the distribution of well-being that would be realized in the ideally just society. On their view, we pick out this ideal distribution by first determining the greatest total sum of well-being that is feasible, and then imagining that this well-being is distributed equally across the population (regardless of whether distributing it in this way would itself be feasible). We then measure deviations from this ideal according to a rule that meets four desiderata: (i) it must rank the ideal distribution as at least as just as all feasible societies, (ii) it must rank one distribution as more just than another if the former is Pareto superior to the latter (better for some and worse for none), (iii) it must sometimes disagree with the utilitarian criterion, and (iv) it “must not depart too far from the difference principle”(discussed below). Finally, we judge that one society is less just than another if it deviates further from the ideal according to this deviation measure.<sup>53</sup>

Though many deviation measures might meet these desiderata, Christiano and Braynen suggest as a “working hypothesis” that the deviation between society  $x$  and the ideal society is given by the following formula, where  $i_1$  through  $i_n$  represent the (equal) well-being levels of individuals 1 through  $n$  in the ideal society, and  $x_1$  through  $x_n$  represent the (perhaps unequal) well-being levels of individuals 1 through  $n$  in society  $x$ :

$$(1/x_1 + 1/x_2 + \dots + 1/x_n) - (1/i_1 + 1/i_2 + \dots + 1/i_n)^{54}$$

In other words, this deviation measure says that that a society deviates from the ideal to the extent that the total sum of individuals’ well-being reciprocals differs from the total sum of well-being reciprocals at the ideal (where if my well-being level is  $a$ , then my “well-being reciprocal” is  $1/a$ ). It is analytic that this deviation measure satisfies Christiano and Braynen’s

---

<sup>53</sup> Christiano and Braynen 2008, p. 408.

<sup>54</sup> Ibid., p. 415.

first three desiderata, and it arguably satisfies their more impressionistic fourth desideratum as well. But does it make essential reference to the ideal? To see that it doesn't, note that if we plug Christiano and Braynen's deviation measure into the criterion that  $x$  is more just than  $y$  when it deviates less from the ideal, we get:

Society  $x$  is more just than society  $y$  if and only if:

$$(1/x_1 + 1/x_2 + \dots + 1/x_n) - (1/i_1 + 1/i_2 + \dots + 1/i_n) <$$

$$(1/y_1 + 1/y_2 + \dots + 1/y_n) - (1/i_1 + 1/i_2 + \dots + 1/i_n)$$

And since the part of this criterion referring to the well-being reciprocals of individuals in the ideal—namely,  $(1/i_1 + 1/i_2 + \dots + 1/i_n)$ —cancels out from each side of the above inequality, this is mathematically equivalent to the following essentially comparative criterion:

Society  $x$  is more just than society  $y$  if and only if:

$$(1/x_1 + 1/x_2 + \dots + 1/x_n) < (1/y_1 + 1/y_2 + \dots + 1/y_n)$$

So Christiano and Braynen's criterion need not be formulated in ideal deviation form after all, but is equivalent to the essentially comparative criterion that one society is more just than another when the total sum of well-being reciprocals in the former society is less than the total sum of well-being reciprocals in the latter. Regardless of whether one thinks this is a plausible criterion—perhaps one that suitably balances a concern with the efficient production of well-being against a concern with its equal distribution—it clearly makes no essential reference to the ideal. In exactly the same way as the utilitarian ideal deviation criterion we considered, it makes the ideal redundant.

Let us turn, next, to Rawls's own rather more complicated conception of justice. For Rawls, the task of ideal theory is to come up with a set of principles that would regulate the ideally just society's "basic structure" or major social and political institutions. In particular, Rawls argues that—given a number of further constraints, including that the principles are

public and (nearly) universally complied with, that the society enjoys a moderate scarcity of resources, and that people have a sense of justice and so are willing to cooperate on terms they view as fair on the condition that others do the same<sup>55</sup>—the ideally just society would realize the following principles:

Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all (*Equal Liberty*).

Social and economic inequalities are to be arranged so that they are (a) attached to offices and positions open to all under conditions of fair equality of opportunity (*Fair Equality of Opportunity*), and (b) to the greatest benefit of the least advantaged (*Difference Principle*).<sup>56</sup>

Rawls adds that Equal Liberty has lexical priority over Fair Equality of Opportunity, which in turn has priority over the Difference Principle, where a lexically posterior “principle does not come into play until those previous to it are either fully met or do not apply.”<sup>57</sup>

So, in Rawls’s ideal society, everyone shares in the most extensive possible equal scheme of basic liberties and enjoys fair equality of opportunity, while remaining inequalities in social and economic benefits redound to the greatest benefit of the worst off. How, though, are we to measure deviations from this society? While Rawls’s official position is that “[t]he measure of departures is left importantly to intuition,” he actually provides us

---

<sup>55</sup> Rawls 1999a, pp. 126-127, *et passim*.

<sup>56</sup> Rawls 1999a, p. 266. I paraphrase these principles very slightly, including by ignoring Rawls’s reference to his “just savings principle” that does not affect my argument. I also ignore an amendment Rawls 2001 makes to the first principle that is similarly irrelevant in the present context.

<sup>57</sup> *Ibid.*, p. 38.

with significantly more guidance than this.<sup>58</sup> For, according to Rawls, “[t]he lexical ranking of the principles specifies which elements of the ideal are relatively more urgent, and the priority rules this ordering suggests are to be applied to nonideal cases as well.”<sup>59</sup> This lexical ranking entails that when comparing societies that fail to satisfy Rawls’s principles, a deviation from Equal Liberty is always more significant than a deviation from Fair Equality of Opportunity, and that each is always more significant than a deviation from the Difference Principle. And to help us measure the extent of each type of deviation, Rawls suggests the following “priority rules”:

Equal Liberty: (a) a less extensive liberty must strengthen the total system of liberties shared by all; (b) a less than equal liberty must be acceptable to those with the lesser liberty.

Fair Equality of Opportunity: an inequality of opportunity must enhance the opportunities of those with the lesser opportunity.<sup>60</sup>

So, as comes out in Rawls’s exposition of these rules, a society that fails to fully realize Equal Liberty deviates from ideal justice less than another if it either (a) provides more extensive equal basic liberties to all or (b) provides more extensive basic liberties to those with less. A society deviates less with respect to Fair Equality of Opportunity if it provides more opportunities to those with less opportunity. And though Rawls never explicitly states a priority rule for the difference principle, he is also clear that a society deviates less with

---

<sup>58</sup> Ibid., p. 216.

<sup>59</sup> Ibid.

<sup>60</sup> Ibid., p. 266. Rawls also includes a priority rule asserting the priority of justice over efficiency, and another referring to his “just savings principle” which I again set aside here.

respect to the difference principle to the extent that it is better for the least advantaged.<sup>61</sup>

Putting Rawls's lexical ranking and priority rules together, then, we come to the following theory of comparative justice. One society is more just than another if it provides a more extensive equal scheme of basic liberties to everyone, or, if basic liberties are unequal, if it provides more to those with less. When basic liberties are not at stake, a society is more just than another if it provides more opportunities to those with less. And when neither basic liberties nor opportunities are at stake, a society is more just than another if it provides greater social and economic benefits to those with less. Though this is a more complicated theory of comparative justice than the last two we considered, it is again one that makes no essential reference to the ideal. So even Rawls falls prey to the redundancy objection. The criteria of comparative justice that he lays out (though never explicitly formulates) are equivalent to essentially comparative criteria, and they therefore render the ideal redundant.<sup>62</sup>

---

<sup>61</sup> For example, when contrasting his position with a view on which we must intuitively balance various “commonsense precepts or objectives of policy” against one another, he notes that the difference principle “*ranks* all combinations of objectives according to how well they promote the prospects of the least favored” (Ibid., pp. 279-280, emphasis mine).

<sup>62</sup> One complication here is that, at least as I read Rawls, his theory of comparative justice—like his theory of ideal justice—is only intended to apply in “reasonably favorable conditions” (that is, to societies that avoid extreme economic, technological, or cultural deprivation). Although questions of justice arise for societies outside of these circumstances, “[t]he principles and their lexical order were not acknowledged with these situations in mind and so it is possible that they no longer hold” (Ibid., p. 216). For further discussion of this point, and of Rawls's comparative theory of justice more generally, see Barrett 2020a.

In all three cases, then, we find the same pattern mentioned earlier. An ideal benchmark theorist begins with the view that a society is unjust to the extent that it deviates from ideal justice, but in the process of working out how to measure deviations from the ideal ends up defending an essentially comparative criterion (or set of criteria) that makes the ideal redundant. And other attempts in the literature, even by those who are aware of Sen's criticisms of the ideal benchmark view, meet a similar fate. For example, Philip Pettit defends a principle of "equal freedom as non-domination" on which ideal justice is achieved when everyone in society is non-dominated—in roughly the sense that they can live as they please rather than "by the leave" of others who have power over them—and holds that societies are unjust to the extent that they deviate from this ideal.<sup>63</sup> Citing Sen, Pettit notes that such deviations can occur along a number of dimensions.<sup>64</sup> But when he actually spells out the ideal deviation criteria that result from fleshing out these dimensions, they are all equivalent to essentially comparative ones: people may enjoy freedom as domination over more or less important choices, these freedoms may be better or worse "entrenched," and so on.<sup>65</sup> Indeed, despite Pettit's adoption of the ideal benchmark view, he appears to admit that, when it comes to evaluating nonideal societies, no reference to the ideal is needed. As he and his co-author, José Luis Martí, write elsewhere: "In any context the ideal will argue for that improvement, assuming there is one available, that makes for a higher degree of nondomination overall."<sup>66</sup> And verdicts about which societies provide a higher degree of nondomination overall make no reference to the ideal society.

---

<sup>63</sup> Pettit 2012, chs. 1-2.

<sup>64</sup> *Ibid.*, pp. 124-125.

<sup>65</sup> *Ibid.*

<sup>66</sup> Martí and Pettit 2010, p. 158.

Of course, these examples do not demonstrate that no plausible criterion of comparative justice makes essential reference to the ideal. But they do provide the beginnings of an inductive argument that suggests that it would be quite surprising if the ideal benchmark view turned out to be true. Not only are many well-known theories of justice essentially comparative (for example, utilitarianism, perfectionism, and various forms of pluralism), but the most explicit attempt (Christiano and Braynen's) and the best-known attempt (Rawls's) to defend ideal deviation criteria that make essential reference to the ideal fall prey to the redundancy objection, as does an attempt by a theorist who explicitly aims to accommodate Sen's worries (Pettit). And, try as I might, I have been unable to find a single example of an ideal benchmark theorist in the literature who articulates an ideal deviation criterion that does not similarly turn out to be equivalent to an essentially comparative one, thus making the ideal redundant. Still, it is worth considering whether all plausible ideal deviation criteria are doomed to redundancy, or whether there is some principled reason why we should expect this of any plausible ideal deviation criterion. Are the examples I have chosen representative? Or have I been ignoring a class of ideal deviation criteria that would genuinely vindicate the ideal benchmark view?

### *Generalizing the Redundancy Objection*

In the remainder of this section, I will complete my argument against the ideal benchmark view by arguing that while there are indeed some logically possible ideal deviation criteria that avoid the redundancy objection, all such criteria should be rejected as implausible on grounds of their explanatory deficiency. In particular, I will show that any ideal deviation criterion that is not equivalent to an essentially comparative one implies the existence of a brute "sweet spot" at which continuing to change a society in a way that, until that point,

makes things more just, suddenly and entirely inexplicably ceases to have any positive effect on justice. Or, at the very least, I will show that ideal benchmark theorists face a heavy explanatory burden here, which none, so far, have even tried to meet.

To see first the claim about logical possibility, take any dimension  $d$  along which we can arrange societies from less  $d$  to more  $d$ . There are two possible essentially comparative criteria that make use of this dimension: one claiming that a society is more just whenever it is more  $d$ , and another claiming that a society is more just whenever it is less  $d$ . The only way for an ideal benchmark theorist to defend an ideal deviation criterion that employs this dimension but that is not equivalent to one of these essentially comparative criteria is therefore to deny both of these claims. Instead, our ideal benchmark theorist must endorse what I will call a *sweet spot criterion* on which making a society more  $d$  (or less  $d$ ) only makes things more just up until a certain sweet spot, after which continuing to make society more  $d$  (or less  $d$ ) either begins to make things less just or stops making any difference to justice.

For example, if—just to take a simple and familiar case—the relevant dimension  $d$  is one that arranges societies from less to more total well-being, then an ideal benchmark theorist may say that the total level of well-being in the ideal society serves as a sweet spot, such that, in measuring deviations from the ideal, a society can deviate by either having too much or too little well-being: if the ideal has  $n$  well-being, then a society with  $n-1$  units of well-being deviates just as much from the ideal as one with  $n+1$  units. Alternatively, an ideal benchmark theorist may claim instead that increasing well-being past the ideal, sweet spot value of  $n$  ceases to have any impact on justice, such that a society with  $n+1$  units of well-being is no more just than one with  $n$ . Unlike the utilitarian criterion, which is able to yield comparative evaluations simply by referring to the total well-being in each of two societies and asking which is greater, such sweet spot criteria really do make essential reference to the

ideal: without referring to the level of well-being realized at the ideal, they cannot yield verdicts about which of two nonideal societies is more just simply by looking at each society's total well-being. After all, if we were simply to keep on increasing total well-being, we would not keep on improving justice, but would eventually overshoot the ideal value. And where, precisely, this overshooting occurs cannot be determined without referring to the level of well-being in the ideal society—that is, without referring to sweet spot value of  $n$ .

I have defined sweet spot criteria in such a way that they represent the only logically possible sort of ideal deviation criteria that are not equivalent to essentially comparative criteria, and therefore the only logically possible ideal deviation criteria that avoid the redundancy objection and so vindicate the ideal benchmark view. But, this vindication comes at a cost: sweet spot criteria are also highly implausible. The problem is not that they posit the existence of a sweet spot, but that they posit the existence of a *brute* sweet spot for which no explanation at all can be given. If increasing total well-being (or more generally, increasing  $d$ ) only improves justice up until a particular sweet spot,  $n$ , after which further increases in well-being (or, more generally, in  $d$ ) decrease justice or at least no longer improve it, then there must be some explanation of this. But, as I will now argue, ideal benchmark theorists can provide such an explanation only by reopening themselves to the redundancy objection. So they can avoid the charge of redundancy only by running into the charge of explanatory deficiency.

In general, there are three ways an ideal benchmark theorist might try to explain the existence of a sweet spot in some justice-relevant dimension—say, why  $n$  is the ideal or sweet spot value of well-being. The first is that, thanks to certain feasibility constraints (for example, certain “fixed constraints of human life” pertaining to human motivation and the

fact that we always face a scarcity of resources)<sup>67</sup> it is impossible to have more than  $n$  well-being in our society. Or, if we take another dimension such as one referring to the total level of oppression in society, we might claim that precisely zero oppression is ideal because we here run into a feasibility constraint of logical possibility: it is logically impossible to have less than zero oppression. This first explanation is not, however, available to an ideal benchmark theorist, because if a sweet spot criterion is to avoid making the ideal redundant, it cannot merely claim that continuing to increase well-being past  $n$ , or to decrease oppression past zero, fails to increase justice because doing so is infeasible, but must rather claim that continuing to increase well-being or decrease oppression past some point ceases to make things more just even when this is feasible. Otherwise, such a criterion ends up equivalent to an essentially comparative criterion: it simply claims that, whenever it is feasible, increasing well-being (or, more generally, increasing  $d$ ) or decreasing oppression (or, more generally increasing  $d$ ) always makes things more just, such that there is no need to refer to the ideal when making comparative evaluations of justice.

A second explanation of why a sweet spot might exist in some justice-relevant dimension is available to those ideal benchmark theorists who adopt a pluralistic theory of justice on which multiple criteria are relevant to our all-things-considered evaluations of justice. For example, such theorists may claim that  $n$  is the ideal value of well-being because, taking into account other criteria of justice—say, one again concerned with oppression—the society that best balances these criteria (given certain feasibility constraints) would have  $n$  well-being. But while this is a perfectly fine explanation of why a sweet spot might exist, it again renders an ideal benchmark theorist vulnerable to the redundancy objection. For even if one adopts a pluralistic view on which other criteria are relevant to justice besides the

---

<sup>67</sup> The term and examples come from Rawls 1999a, 216.

sweet spot criterion in question, one is still committed to thinking that, *holding all other criteria fixed*, deviating further from the ideal in the dimension picked out by one's sweet spot criterion makes things less just. A sweet spot criterion that does not make the ideal redundant must therefore claim that even when it is feasible to have more well-being than  $n$  without this increase in well-being conflicting with any other criterion, increasing a society's total well-being past  $n$  ceases to make things more just. Otherwise, such a criterion again proves equivalent to an essentially comparative criterion: it simply claims that, holding all other criteria fixed, increasing well-being (or, more generally, increasing  $d$ ) always makes things more just when such increases are feasible.

A third explanation of why a sweet spot may exist in one dimension is that this is explained by some more fundamental criterion (or criteria) of justice concerning some more fundamentally justice-relevant dimension (or dimensions). If, for example, there is an ideal tax rate such that either raising or lowering this tax rate would make things less just, then this is most plausibly explained by reference to other criteria: perhaps the sweet spot with respect to our tax rate is explained by its maximization of well-being, or by its ability to best balance efficiency against equality given certain feasibility constraints. Or, to take another example, suppose that improving the income level of a society's poorest member increases a society's justice until she hits a certain income level, after which further increases have no effect on justice. This may be explained by the fact that justice is not fundamentally concerned with the income level of the poorest, but rather with some more fundamental criterion that explains the relevance of the poorest's income level—say, one concerned with reducing poverty, such that once the poorest member hits the poverty line further increases in her income no longer have any impact on justice since they no longer reduce poverty. But this explanation will not help the ideal benchmark theorist, in the first place, because our earlier

discussion of Sen's critique has revealed that ideal deviation criteria must refer to dimensions that are fundamentally relevant to justice, since they otherwise run into second-best problems and so the charge of implausibility. And, even setting this aside, explaining the existence of a sweet spot in one dimension by reference to other more fundamental criteria should provide no solace to the ideal benchmark theorist, since it implies that we don't really need to invoke the sweet spot criterion in question in order to make or explain the truth of comparative evaluations of justice: we can simply appeal to the more fundamental criteria that explain the existence of a sweet spot in our less fundamental criterion. So this sort of explanation is not available to the ideal benchmark theorist either. It does nothing to explain the existence of a sweet spot in our fundamental criteria of justice, and so nothing to rebut the charge that our fundamental criteria make the ideal redundant.

We have now examined three ways how, for some given dimension  $d$ , increasing  $d$  might make things more just up until some sweet spot, after which further increases no longer generate improvements. First, it may be infeasible to make a society more  $d$ ; second, it may be infeasible to make a society more  $d$  without also setting back the satisfaction of some other criterion; and third, it may be that  $d$  is not actually a dimension that is fundamentally relevant to justice, but is rather one whose relevance is explained by some more fundamental criterion. But we have also seen that to endorse any of these three explanations is to abandon a sweet spot criterion that makes essential reference to the ideal in favor of one that is equivalent to an essentially comparative criterion on which, for some fundamentally justice-relevant dimension  $d$ , increasing  $d$  always makes things more just when it is feasible to do so without setting back the satisfaction of some other criterion. So, in order for an ideal deviation criterion to avoid making the ideal redundant, it must claim that, for some justice-relevant dimension  $d$ , making a society more  $d$  makes it more just until it hits some sweet

spot level where making a society more *d* suddenly ceases to make it more just—where the existence and position of this sweet spot is not explained by any feasibility constraint, by any conflict with other criteria, or by any more fundamental criterion. Unless an ideal benchmark theorist can come up with some fourth explanation for the existence and position of a sweet spot, she must therefore claim that it is simply a brute, inexplicable fact that ranking higher in some dimension that is fundamentally relevant to justice makes things more just until, suddenly, it doesn't. But it is difficult to imagine what this explanation could be. And though I cannot rule out its possibility, it suffices to say that no ideal benchmark theorist has ever even tried to provide one, and that until a benchmark theorist proves otherwise, it therefore seems safe to assume, from this point forward, that no such explanation is forthcoming.

Our earlier observation that all ideal deviation criteria articulated by actual benchmark theorists make the ideal redundant should therefore come as no surprise. None commit themselves to a criterion as one that contains a brute, inexplicable sweet spot. Indeed, in order to head off a possible objection, it is worth noting that even those theorists that appear to defend sweet spots in other normative domains take pains to point out that such sweet spots are not brute facts, but are rather explained in one of the three ways we have considered. For example, though Aristotle famously claims that virtue involves a mean between excess and deficiency—such that courage, for example, involves being neither too fearful nor not fearful enough—even he explains such sweet spots by reference to a more fundamental criterion of rationality or practical wisdom. In the case of courage, practical wisdom determines that there are certain things one should fear (to a particular degree), and others one should not fear, such that failing to hit the mean of courage involves fearing those things one should not fear, or not fearing those things one should fear (to the right

degree).<sup>68</sup> So it is this prior standard of what one should and should not fear (and to what degree) that explains what counts as either fearing too much or too little, rather than a brute fact about fear being good up until a sweet spot, and then bad. Thus, even Aristotle does not posit a *brute* sweet spot, and this is what allows him to retain whatever plausibility his view has. As he himself puts it: though virtue is a “mean” in one dimension, it is an “extreme” in the dimension of practical wisdom, and its extreme position in the latter dimension is what explains the sweet spot in the former.<sup>69</sup>

So criteria of comparative evaluation that posit brute sweet spots are implausible on grounds of their explanatorily deficiency, and must therefore be rejected. But, as we have seen, the only alternative to such sweet spot criteria are those that are equivalent to essentially comparative criteria, and that therefore make the ideal redundant. So the ideal benchmark view is defeated. Criteria of comparative justice do not make essential reference to the ideal. Even though we may articulate them in ideal deviation form, all plausible criteria turn out to be equivalent to essentially comparative criteria, capable of yielding verdicts about when one society is more just than another simply by referring to the features of those two societies. And so all plausible criteria of comparative justice render the ideal redundant.

## 2.5. Facilitation and Obstruction

To this point, I have been concerned with rebutting the ideal benchmark view, according to which criteria of comparative justice make essential reference to the ideal. This was important to do because, otherwise, ideal theory would have a strong claim to theoretical priority over problem solving, in the sense that we would first need to work out the former

<sup>68</sup> Aristotle 2000, 1006b.

<sup>69</sup> Aristotle 2000, 1107a. Compare Annas 1993, pp. 59-61.

before engaging in the latter. For, as we noted earlier, if problem solving requires us to invoke criteria of comparative justice—as I have granted—and if criteria of comparative justice make essential reference to the ideal—as the ideal benchmark view claims—then it follows that problem solving requires us to refer to the ideal as well. Now that we have seen that plausible criteria of comparative justice do not make essential reference to the ideal, however, this claim of priority falls by the wayside. But there remains the further possibility that ideal theory might facilitate problem solving without being prior to it, and that it might therefore remain practically relevant to reform theory in this way.

Ideal theorists have pointed to three ways such facilitation might occur. First, ideal theory might help us to develop our criteria of comparative justice through something like a process of reflective equilibrium.<sup>70</sup> For example, suppose we develop a conception of an ideal society in which race has no influence on one's treatment or life prospects and laws are therefore color-blind, making no reference to race at all. This is certainly an important data point when working out our criteria of comparative justice since it at least places one constraint on them: whatever else is true of them, they had better be able to explain why a color-blind society is ideal.

But, second, since the ultimate goal of problem solving is not merely to develop abstract criteria, but to apply them by working out solutions to problems or concrete ways of making our society more just, ideal theory might also facilitate problem solving by helping us to expand our understanding of the space of social possibilities, thereby helping us to come up with new and creative solutions.<sup>71</sup> For example, in thinking through a market socialist ideal in which a free market sets individuals' pre-tax income, redistributive taxation equalizes

---

<sup>70</sup> Gilabert 2012, Stemplowska 2008, Swift 2008.

<sup>71</sup> Gilabert 2012.

post-tax income, yet individuals remain motivated to work by an egalitarian “ethos,” we may come to recognize the possibility of using similar ethos-based solutions to solve more immediate problems that we have historically attempted to solve via material incentives.<sup>72</sup>

And third, since the concrete problems we face—the ways our society is less just than it could be—are themselves often hidden from view or difficult to understand, ideal theory might help us to identify or appreciate such problems, thereby guarding against a complacent acceptance of the status quo.<sup>73</sup> If, for example, we spell out a conception of an ideal society free of gender oppression, then comparing this society to our actual society may help us notice the ways our current gender relations fall short: perhaps in the ideal society individuals with different genders would not only enjoy formal equality under the law, but equality with respect, say, to patterns of economic dependence, and realizing this can help open our eyes to such problematic patterns in our own world.

These three facilitating roles of ideal theory—its ability to help us to develop our abstract criteria, to come up with novel solutions, and to appreciate existing problems—may seem uncontroversial enough. But against each role, critics have warned that ideal theory may obstruct problem solving more than it facilitates. Against the first role, they argue that focusing on the ideal may distort our theorizing about criteria of comparative justice, in much the same way that attempting to fit a scientific theory to a single data point may sometimes lead to a distorted theory.<sup>74</sup> For example, though a color-blind ideal may indeed put a constraint on our criteria of comparative justice—they must be able to explain why this is ideal—there are many different criteria that can meet this constraint, and focusing too

---

<sup>72</sup> Compare Carens 1981, Cohen 2008.

<sup>73</sup> Gilabert 2008, Stemplowska 2008.

<sup>74</sup> Compare Anderson 2011, ch. 1, Goodin 1995.

much on the ideal may lead us to endorse one that most simply explains why this is ideal rather than one that best fits our convictions about a variety of actual and hypothetical nonideal societies as well. In the case at hand, we may jump to the conclusion that we should accept a criterion of comparative justice on which color-blind policies are always more just than color-conscious policies. But it might turn out that had we thought carefully about a variety of actual or hypothetical nonideal cases, we would have come to realize that the actual criterion that explains why a color-blind society is ideal (if it is) also explains why color-conscious policies are more just than color-blind ones in some nonideal circumstances (if they are)—say, because this criterion is concerned with equality of opportunity.<sup>75</sup>

Second, though theorizing about the ideal is certainly one way to expand our understanding of the space of social possibilities, critics argue that ideal theory is liable to focus us on less relevant regions of this space: on unrealistic rather than viable solutions to the problems we face.<sup>76</sup> For example, though market socialism is certainly interesting, the history of attempts to realize other versions of socialism or to structure an economy without relying heavily on material incentives should certainly give us pause, and make us turn instead to the various ways that actual societies have successfully dealt with their problems. More generally, we often learn more about what solutions are available to us by beginning with a careful analysis of the causal workings of the problems we face and by brainstorming various ways we might fix them, or by engaging in cross-historical and -societal surveys of how other past and present societies have tried to solve similar problems, than by trying to imagine an ideal society free of such problems altogether. And in this regard, ideal theory may serve more to distract than to facilitate. We may explore more relevant regions of the

---

<sup>75</sup> See Anderson 2011, ch. 1.

<sup>76</sup> Compare Anderson 2011, ch. 1, Schmidtz 2011, Wiens 2012.

space of social possibilities by directly attempting to figure out ways of solving our problems and exploring the regions of this space that correspond to social actualities than by attempting to figure out which point in this space would be best.

Third, though it is again possible that comparing an ideal society to our own will help us notice or better understand ongoing problems, critics suggest that, in practice, focusing on the ideal is more likely to blind us to such problems than to reveal them. In the first place, we often develop our conception of the ideal in response to problems that we have already identified rather than the other way around. Indeed, some argue that this is not only typically but necessarily true of our moral and political thinking.<sup>77</sup> But even if a conception of the ideal can sometimes help us to appreciate problems in our current society, it strains plausibility to claim that turning away from our actual, problem-ridden society and focusing instead on trying to identify an ideal, problem-free society is generally a better way to identify and understand these problems. For example, if the ideal would contain no oppression, then ideal theory will involve very little theorizing about the actual workings of oppressive arrangements, and so will ill-equip us to identify or understand the subtle and insidious forms of oppression that pervade actual societies. The most famous proponent of this sort of argument, Charles Mills, has gone so far as to argue that the disciplinary dominance of ideal theory therefore serves as an *ideology* “in the pejorative sense of a set of group ideas that reflect, and contribute to perpetuating, illicit group privilege.”<sup>78</sup> Writes Mills:

Can it possibly serve the interests of *women* to ignore female subordination...?

Obviously not. Can it possibly serve the interests of *people of color* to ignore the centuries of white supremacy...? Obviously not. Can it possibly serve the interests of

---

<sup>77</sup> See Anderson 2011, ch. 1, who follows Dewey 1922, esp. pp. 260-261.

<sup>78</sup> Mills 2005, p. 166.

the *poor and the working class* to ignore the ways in which an increasingly inequitable class society imposes economic constraints that limit their nominal freedoms, and undermine their formal equality before the law? Obviously not. If we ask the simple, classic question of *cui bono?* then it is obvious that ideal theory can only serve the interests of the privileged.<sup>79</sup>

In other words, Mills argues that when we engage in ideal theory and imagine a society free from oppression and other instances of injustice, we are “abstracting away from realities crucial to our comprehension of the actual workings of injustice in human interactions and social institutions, and thereby guaranteeing that the ideal... will never be achieved.”<sup>80</sup> He therefore thinks it “no accident that historically subordinated groups have always been deeply skeptical of ideal theory, generally see its glittering ideals as remote and unhelpful,” and typically prefer to “make theoretically central the existence and functioning of the actual nonideal structures” that produce the problems they face.<sup>81</sup>

Mills overstates his case. It is simply not plausible that the dominance of ideal theory really *guarantees* that we will never reach the ideal. But the weaker claim that the widespread practice of ideal theory—at least insofar as it is practiced to the exclusion of other forms of theorizing about justice—hinders the pursuit of greater justice, or at least our theorizing about such, has more bite. And the same is true of the other two criticisms as well. In each case, critics go too far if they deny that ideal theory ever helps us to develop our criteria, to come up with new solutions, or to identify and understand ongoing problems. Ideal theory can certainly play these roles. Yet at the same time, critics identify genuine limits of ideal

---

<sup>79</sup> Ibid., p. 172.

<sup>80</sup> Ibid., p. 170.

<sup>81</sup> Ibid.

theory that warn against an overreliance on it. Our judgments about the ideal serve as relevant data points in developing our principles of comparative justice, to be sure, but we risk overfitting our theory to this data if we fail to investigate actual or hypothetical nonideal societies that provide equally important data. Ideal theory can help us to explore the space of social possibilities by examining its outer limits, but this is no substitute for engaging in careful empirical analysis of the problems we face and of actual attempts to solve similar problems. And though a conception of the ideal can sometimes help us to guard against complacency and to notice new problems, the risk of complacency also arises for those who spend their times gazing toward the ideal rather than actually examining and gathering evidence about the problems we face—a process that, Mills and others emphasize, includes listening carefully to what people have to say about their problems rather than just speculating from the armchair.

Now, it is difficult to articulate precisely to what extent ideal theory is useful, or at what point relying on it becomes problematic. But the limited relevance of ideal theory to problem solving nevertheless comes out clearly when we note that it has little more claim to such relevance than theorizing about dystopia. Much like a conception of the ideal, a conception of dystopia may help us to develop our criteria of comparative justice by providing us with new data points, or to explore the space of social possibilities by alerting us to ways that “solutions” might work out horribly—just think of how adeptly Kurt Vonnegut’s “Harrison Bergeron” illustrates the folly of criteria that license attempts to resolve inequality through leveling down, or how George Orwell’s *1984* does the same for totalitarianism. And theorizing about dystopia can similarly help us to notice problems in our current society through alerting us to resemblances rather than gaps—here, Margaret Atwood’s *Handmaids Tale* stands out for its ability to alert us to gender inequalities in our

current society through its depiction of a dystopian society in which such inequalities are vastly exaggerated yet still alarmingly recognizable. But despite its ability to sometimes play these three helpful functions, no one would claim that dystopian theorizing plays, or should play, a central role in problem solving. And ideal theorists provide us with no grounds for thinking that we should view ideal theory any differently.

Ideal theory thus plays a modest, ancillary role in problem solving. It enjoys no priority over problem solving, and working out a conception of the ideal is not a central task of problem solving, but rather something that we might sometimes find helpful along the way—just like it may sometimes be helpful to theorize about dystopia. If ideal theory deserves pride of place in reform theory, it must therefore be because reform theory concerns more than just problem solving, as proponents of the ideal target view claim. So let us turn to that possibility now.

## CHAPTER 3 – CHASING TARGETS

## 3.1 Short-term versus Long-term Justice

Whereas ideal theorists begin by attempting to describe the ideally just society, problem solvers begin with the problems of our own. They aim to identify actual instances of injustice in our society, to diagnose them by tracing out their causes, and to propose narrowly targeted solutions. For example, a problem solver may begin with a particular instance of status or resource inequality, of discrimination or oppression, of a human rights violation or an unjustifiable restriction on freedom. And she might work to uncover its causes—the institutions and other concrete features that perpetuate or sustain it—and to come up with a targeted solution that would mitigate or eliminate it: say, a change to our laws or social norms. Problem solvers therefore adopt a relatively narrow and short-term perspective when theorizing about social reform. They aim to identify “remediable injustices” in our society along with promising remedies to them—remedies that would promote justice by chipping away at the many problems of injustice that confront us.<sup>82</sup>

Ideal theorists who endorse the *ideal target view* offer another approach. They begin not with the injustices we currently face, but by attempting to outline what the ideally just society would be like, before then figuring out, as Rawls puts it, “how this long-term goal might be achieved, or worked toward, usually in gradual steps.”<sup>83</sup> So instead of diagnosing actual problems of injustice, they aim to identify the most just, problem-free society we could ever achieve. And instead of working out which changes to our current society would

---

<sup>82</sup> The term “remediable injustice” is, again from Sen 2009, p. vii, but in describing the problem solving approach I hew most closely to Wiens 2012. See also Anderson 2010, ch. 1 and Schmitz 2011.

<sup>83</sup> Rawls 1999b, p. 89.

mitigate or eliminate present injustice, they ask which changes would constitute progress toward this ideal. Ideal target theorists therefore adopt a more comprehensive perspective than problem solvers—one concerned with the ideally just society as a whole rather than with targeted solutions to specific problems. And their perspective is also longer-term, concerned not with promoting justice in the short term but, most centrally, with identifying a long-term goal—an end-goal—“to guide the course of social reform.”<sup>84</sup>

When we compare these two approaches side by side, ideal target theorists might seem alarmingly farsighted. It would be one thing if we needed to engage in ideal theory before we could get down to any serious problem solving, as ideal benchmark theorists claim. But we have seen (in Chapter 2) that the ideal benchmark view must be rejected. So why would anyone think that political philosophers should focus on identifying a far-off goal of ideal justice, rather than on diagnosing and solving the problems of injustice that we currently face? Isn't the whole point of theorizing about justice to work out ways of making our current society more just? And, if so, shouldn't we engage in ideal theory only to the limited extent that it facilitates problem solving, in the ways we have sketched in the last chapter?

One way ideal target theorists might respond to this charge of hyperopia is by falling back on the anti-practicalist defense of ideal theory (discussed in Chapter 1), on which the value of ideal theory depends on something other than its practical relevance. Even if an understanding of the ideally just society provides us with little guidance about how to reform our actual society, ideal theory might still be worth doing, say, because it is important to understand what the ideally just society would be like. But I will continue to set aside this response for now, and focus instead on how ideal target theorists have attempted to rebut

---

<sup>84</sup> Rawls 1999a, p. 215.

the charge that ideal theory is practically irrelevant. In particular, ideal target theorists often claim that an exclusive focus on problem solving is itself too myopic—that even though we may solve particular problems of injustice without any ideal in mind, implementing such solutions is not, by itself, a reliable way to promote overall long-term justice, since doing so may, as A. John Simmons puts it, “retard,” “stall,” or “permanently block... movement toward overall justice.”<sup>85</sup> So when evaluating potential reforms, we cannot focus only on their short-term effects on particular problems of injustice, but must balance this against their promotion of greater justice in the long term. And while problem solving is a good way to figure out the first half of this balance, it is not enough. We must also take into account overall long-term justice, and for that, we need ideal theory.

Ideal target theorists have a point. Societies are complex—they are composed of many interacting parts—and, as I shall argue in this chapter, this social complexity entails that ameliorating particular problems of injustice may indeed set back the achievement of greater long-term justice. So reform theory does require more than problem solving, as ideal target theorists rightly argue. But it does not follow that we should do ideal theory. In fact, as I will also argue in this chapter, the very complexity that generates a conflict between ameliorating immediate problems of justice and promoting overall long-term justice also renders ideal target theory epistemically overdemanding for beings like us: it makes it impossible for us either to identify the ideal or to track our progress toward it, at least with sufficient confidence to warrant its pursuit. Furthermore, considerations of social complexity suggest that there isn’t even a fixed ideal for us to pursue in the first place: there are not only epistemic limitations to our ability to pursue the ideal, but practical limitations, since to pursue the ideal is to chase a moving target. So, thanks to complexity, problem solving is

---

<sup>85</sup> Simmons 2010, p. 21. Compare Buchanan 2004, ch. 1, Robeyns 2008, and Valentini 2012.

unsatisfactory and ideal target theory is impracticable. The remaining question is how we ought to theorize about social reform in a complex world, and, in particular, how we should theorize about long-term justice without recourse to ideal theory.

Answering this question will be the focus of the next chapter, where I will argue that rather than attempting to identify ways of making *progress toward* the ideal, reform theorists should instead approach questions of long-term justice by working out how to make our social arrangement more *progressive*: better at getting better, or more conducive to further improvements in general (though not necessarily to the achievement of any antecedently specified goal). But before we turn to a development of this *progressiveness approach*, we must first get clear on why ideal theorists are right to point out the genuine limits of the problem solving approach, but wrong to think that ideal theory provides a viable supplement. In this chapter, I will therefore explain each of these points in turn, before summing up our largely negative conclusions about the practical relevance of ideal theory reached in the last two chapters. Of course, some ideal theorists may find this unsatisfactory, since it leaves untouched the anti-practicalist defense of ideal theory on which ideal theory is worth doing regardless of its practical value. But a proper engagement with this position will have to await our discussion of the progressiveness approach in the next chapter, after which we will be better situated to consider the merits of the two approaches on their own terms.

### 3.2 Complexity and Problem Solving

Let us begin with some terminology. A “social arrangement” as I use the term, refers to a set of formal and informal institutions (for example, laws, policies, and social norms), as well as the background conditions that are causally relevant to their functioning (for example, facts about the natural environment, demographics, technology, human psychology, and people’s

empirical and normative beliefs). Both institutions and background conditions are what I call “social features”: they serve as the inputs to our (often implicit) models of how a social arrangement produces its effects. The features represented by the outputs of such models I call “outcomes,” and by the “justice” of a social arrangement I mean an overall evaluation of the outcome it produces given some all-things-considered criterion of comparative justice. This may be an external criterion of the sort I discussed in the last chapter, which takes into account and balances one or more criteria relating to basic principles or values like freedom, equality, oppression, and procedural justice. Or it may be an internal criterion concerned with what can be justified to actual individuals—holding, for example, that the justice of an arrangement depends on how highly it ranks on the criteria of those living under it, rather than against independently specified values or principles. Once again, I make no attempt to defend any particular criterion of comparative justice in this work, but instead invite the reader to apply her favored criterion to the issues at hand. I assume only that one does not endorse a criterion that makes essential reference to the ideal (since my argument in the last chapter showed all such criteria to be implausible), and that one is not an “institutional fundamentalist” who denies that the outcome produced by a social arrangement is at all relevant to its justice.<sup>86</sup>

I emphasize the distinction between a social arrangement, the outcome it produces, and its justice to highlight something that, though obvious upon reflection, is too often omitted in philosophical discussions of social reform: that our evaluation of a social arrangement’s justice is always (logically, though perhaps not temporally) a two-step

---

<sup>86</sup> For a critique of such institutional fundamentalism, see Sen 2009, ch. 3, who aims to show that Rawls and various others are committed to it. Sen’s interpretation of these theorists has been widely challenged, but his critique of institutional fundamentalism has not.

procedure.<sup>87</sup> We must first map a social arrangement to an outcome, and only then can we map that outcome to its justice. For example, to determine whether implementing a minimum wage would improve justice, the first step is to ask what outcome this change would produce: how would it affect, say, unemployment, prices, and income inequality? And the second step is to evaluate that outcome given our all-things-considered criterion of justice: would the predicted change in such variables amount to a net increase or decrease of justice, understood, say, in Rawlsian, utilitarian, republican, or libertarian terms? Though perhaps the bulk of political philosophy concerns the appropriate criterion to use at this second step, my focus in this chapter is on the first step where we map social arrangements to outcomes. For convenience, I will therefore often speak as if we can skip the second step and map social arrangements to justice directly. But it is important to remember what this involves. It requires us to employ a “predictive model” of how various social features interact to produce an outcome, which we must then evaluate in terms of its justice.<sup>88</sup>

The notion of an interaction is essential to any understanding of complexity. To say that something is complex is not merely to say that it is complicated—it is to say that it has many interacting parts.<sup>89</sup> In analyses of societies as complex systems, these parts are usually

<sup>87</sup> I adapt these distinctions, and the idea of this two-step procedure, from Gaus 2016, ch. 2.

<sup>88</sup> Ibid.

<sup>89</sup> The science of complexity is rapidly growing, and has been applied to a staggering number of phenomena: from physics and biology, to the behavior of flocks of birds and crowds of people, to now, increasingly, economic and public policy issues. Here, I make no attempt to survey this literature, but only to draw out some of its central insights as they apply to questions of social reform. But for helpful discussions of complexity in general see Holland 2016, Mitchell 2011, and Page 2010, and for examples of its recent applications to

thought of as people, whose interactions produce emergent patterns that no individual intended or perhaps even foresaw.<sup>90</sup> The classic example of this is Adam Smith’s discussion of how, given certain social arrangements, the market interactions of individuals each pursuing their own self-interest leads to greater social welfare.<sup>91</sup> But such “invisible hand” processes are not always for the good: market interactions, to take the same example, may also lead to economic inequality, environmental destruction, and financial crises.<sup>92</sup> And Thomas Schelling has shown how, again, given certain arrangements, individuals with mild preferences not to live in neighborhoods in which their own racial group is a small minority can lead, through a mechanism of tipping points and cascades, to stark racial segregation. Similar mechanisms plausibly contribute to segregation and polarization along a number of dimensions (not just race, but, for example, gender and religion) in a variety of domains (not just neighborhoods, but, for example, schools and industries).<sup>93</sup>

An understanding of the various ways that individuals may unwittingly interact to produce both good and bad emergent phenomena in the presence of different social and especially institutional arrangements—which, after all, structure such interactions by 

---

economics, see Arthur 2015, to public policy, see Colander and Kupers 2016, and to political philosophy, see Gaus 2016, 2018, 2019, D’Agostino 2018, Page 2018, and (though less recently) Hayek 1964 and Mill 1974.

<sup>90</sup> For an overview of this sort of complexity, see Miller and Page 2007, Part IV.

<sup>91</sup> Smith 1776.

<sup>92</sup> These three issues are at the heart of recent attempts to rethink economic policy from a complexity perspective, especially in the wake of the 2008 financial crisis. See OECD 2017.

<sup>93</sup> Schelling 1971, 1978. Of course, much actual segregation is produced very intentionally, rather than as the unintended consequence of diffuse individual interactions.

constraining and incentivizing certain forms of behavior—is important to any analysis of social change. In the first place, it dispels us both of the overly rationalistic view that individuals can only produce just outcomes by explicitly aiming to do so, and of the overly complacent view that individuals pursuing their own projects reliably promote justice in all circumstances: both views fail to recognize that the causal relation between the achievement of greater justice and the intentional pursuit of other goals crucially depends on which social arrangements are in place.<sup>94</sup> In the second, it helps to explain why the effects of social change are so difficult to predict, as it is one important source of the unanticipated consequences that often accompany such changes.<sup>95</sup> Going forward, however, my focus will be not on the micro-level complexity that characterizes interactions between individuals, but instead on two forms of what we might call macro-level complexity. It will be on the way that different social features (institutions and background conditions) interact to produce outcomes, as well as to cause or prevent changes in other social features themselves.

Consider, first, *combinatorial complexity*. This is the sort of complexity that arises when predicting what outcome will be realized by the interaction of multiple social features. As is now commonplace among institutional theorists, the operation of any one institution is

---

<sup>94</sup> Compare Wilson 2016, p. 44.

<sup>95</sup> Another source is a lack of technical knowledge or understanding, as when we avoid ground-level pollution by building smoke stacks only to create acid rain, or improve public sanitation to wipe out cholera only to produce a polio epidemic. And a third is simply that to err is human—that even in cases where we have the resources to accurately predict the consequences of a social change, we often make mistakes. I return to related issues shortly, but for a book chock full of these and other examples, see Tenner 1996, and for a classic analysis of the various sources of unanticipated consequences, see Merton 1936.

importantly dependent on the presence and operation of other institutions, as well as on background conditions.<sup>96</sup> Often this phenomenon is discussed by economists under the rubric of the “general theory of second best.” As R. G. Lipsey and K. J. Lancaster famously proved, if market institutions fail to meet the set of “optimality conditions” that ensure a Pareto optimal outcome (in which no one can be made better off without someone else being made worse off), the second-best outcome is not necessarily achieved by satisfying more rather than less of these conditions: market institutions that fail to meet two optimality conditions might be Pareto superior (better for some and worse for none) than those that fail to meet only one.<sup>97</sup> This, however, is just one instance of the more general phenomenon of combinatorial complexity. If we cannot have the optimal or ideal social arrangement, but the features of that arrangement interact to produce an outcome, then satisfying more of the features that compose the ideal does not necessarily result in an improvement. More generally, combinatorial complexity entails that the effect of any two social features cannot be reduced to the sum of the effects of each feature by itself. So even when each of two social changes would, on their own, improve justice, both changes together might not.

Combinatorial complexity is ubiquitous. For example, market institutions only produce efficient outcomes given (among other things) a background of social trust and the absence of norms prohibiting profit seeking.<sup>98</sup> Criminal prohibitions only command respect and compliance in the presence of a social norm of legal obedience, and when laws conflict too sharply with other norms this often gives rise to compliance and enforcement problems—sometimes reinforcing rather than undermining the behavior the law seeks to

---

<sup>96</sup> See especially North 1990.

<sup>97</sup> Lipsey and Lancaster 1956.

<sup>98</sup> This point goes back to Smith 1776, but see Platteau 2000, chs. 5, 7.

abolish.<sup>99</sup> Color-blind policies may seem just in isolation, but may only serve to reinforce racial inequalities produced by other features of our social arrangement; in such cases, color-conscious policies may promote justice, even if they would undermine it given background equality.<sup>100</sup> The same is true more generally of institutions that are designed to correct or compensate for injustice: they may be beneficial in the presence of particular forms of injustice, but detrimental in their absence.<sup>101</sup> Or consider again the possibility that while two changes might each improve justice on their own, the combination of them might not. For example, instituting generous entitlement programs might improve justice, and opening our borders might similarly be an improvement, but doing both together might be disastrous: the influx of immigrants might result in the entitlement programs being stretched beyond the breaking point.<sup>102</sup> Or perhaps instituting any number of entitlement programs on their own would produce justice, but instituting too many would result in massive increases in our deficit and ultimately government default and ruin.<sup>103</sup>

These instances of combinatorial complexity all involve social features interacting to produce outcomes. But social features also interact in the sense that prior social changes may further or set back later social changes, and this gives rise to a type of path-dependency that I will call *transitional complexity*.<sup>104</sup> For example, even if opening our borders and enacting

---

<sup>99</sup> Barrett and Gaus forthcoming.

<sup>100</sup> Anderson 2010, ch. 8.

<sup>101</sup> Estlund 2014, pp. 120-121.

<sup>102</sup> Related worries are common in the literature on immigration. See Carens 2013, ch. 12.

<sup>103</sup> I adapt this example from Gaus 2016, pp. 63-64

<sup>104</sup> A process is path-dependent when what happens earlier in the process affects what happens later. But not all path-dependence is driven by complexity, or interactions between

more generous entitlement programs would indeed improve justice, opening our borders first might make it more difficult to implement these entitlement programs later, if (as some empirical evidence suggests) influxes of immigration undermine public support for such programs.<sup>105</sup> More general phenomena relevant to transitional complexity include lock-in and backlash. Lock-in occurs when a social change prevents further changes, often because it generates interest groups who are able to maintain the new status quo.<sup>106</sup> Backlash occurs when a social change, say, the prohibition of alcohol or the passing of the Fugitive Slave Act, results in countervailing efforts to reverse that change, or perhaps in other changes that undercut the effect of that former change.<sup>107</sup> Or consider the possibility of transitional institutions, whose function is to cause further change before they can fall away in something like the manner of booster rockets—for example, an institution that “incorporate[s] politically disadvantaged groups into the democratic process” by employing mechanisms that “differentially favor these groups” before it can be abolished once political power is equalized.<sup>108</sup> Regardless of whether one thinks such institutions are ultimately

---

different features—for example, path-dependency can also arise simply due to the cost of making changes to a simple system. For a lucid discussion, see Page 2006.

<sup>105</sup> Carens 2013, ch. 12.

<sup>106</sup> For a helpful discussion of lock-in, and the tendency of institutions to produce groups with a stake in maintaining the status quo, see North 1990, who builds on Arthur’s 2015, ch. 4, analysis of technological path-dependency.

<sup>107</sup> I borrow these examples from Stuntz 2000. For a general discussion of such “reactive” as opposed to “self-reinforcing” sequences, see Mahoney 2000. This tracks the distinction between “negative” and “positive” feedback in the literature on complexity.

<sup>108</sup> Knight and Johnson 2011, p. 247

justified, the very coherence of the justification offered for them depends on transitional complexity: the fact that a change to one social feature may cause or prevent changes to others. Perhaps even more obviously, changes to secondary or constitutional rules (rules for changing other rules) clearly interact with future changes to our institutions. For example, in a democracy, different voting schemes predictably lead to different policies, as do different constitutional constraints on what is subject to democratic rule.<sup>109</sup>

To see the importance of combinatorial and transitional complexity, suppose that neither held. In that case, problem solving would be the perfect approach to theorizing about social reform. We could identify instances of injustice, trace their causes to discrete social features, and identify social changes that would not only solve the problem we are focused on, but, in so doing, promote overall long-term justice as well. We could, for example, identify a change to employment legislation to end worker oppression, to the economic system to mitigate income inequality, to healthcare policy to minimize preventable deaths, to educational institutions to clear up status inequalities, to immigration policy to ameliorate global poverty, to the criminal law to mitigate racial inequality, to our gender role norms to decrease gender discrimination, and so on, without ever needing to consider how these different “solutions” would interact. We would not have to worry, for example, that the effect of different employment policies depends on what exit options employees have, which in turn depend on what sorts of economic, healthcare, and other social safety net programs are in place, whose cost and efficacy depend on education policies, which also affects health, criminality, the economy, norms relating to gender and race, and so on, in various crisscrossing ways, throughout the entire network of interactions. (This is

---

<sup>109</sup> Riker 1982 provides an authoritative discussion of the way that different voting rules may produce different results even given the same voter preferences.

combinatorial complexity.) Nor would we have to worry that changing, say, immigration and education policies would affect which further changes would occur—by, for example, generating an anti-immigrant backlash, strengthening teachers’ unions who are now able to lock us into our current system, or changing the makeup of the electorate and so which further changes are likely to be democratically enacted. (This is transitional complexity.) We could, in other words, engage in problem solving without ever having to consider whether the “solutions” we generate would undermine the achievement of overall or long-term justice—either by interacting to make things less just in the short-term or by setting back the achievement of greater justice in the long-term.

But combinatorial and transitional complexity do exist. So, bringing things full circle, this is why ideal theorists are right to point out that reform theorists cannot focus only on identifying social changes that would ameliorate particular problems of injustice: we must also take into account how these changes would interact to produce outcomes, and to change or stabilize other social features.<sup>110</sup> And this suggests that we must expand our vision along two dimensions. First, we must adopt a more comprehensive, holistic attitude to

---

<sup>110</sup> Simmons pp. 21-22, appears to have combinatorial complexity in mind when he argues that “[t]here is no reason to suppose in advance that justice in one domain is independent of justice in other domains,” and transitional complexity in mind when he warns of the potential for backlash. See also Shelby’s 2016, p. 2, *et passim*, critique of what he calls the “medical model of social reform” on the grounds that its attempt to come up with “narrowly targeted and empirically grounded interventions” manifests a form of “status quo bias” in its failure to examine ways in which the basic structure of society “needs fundamental reform.” This critique implicitly assumes the existence of complexity, without which the pursuit of narrowly targeted reforms could never conflict with more comprehensive reforms.

evaluating social arrangements rather than one that focuses only on their component parts. And second, we must adopt a longer-term perspective—one that takes into account not only relatively short-term solutions to particular problems of injustice, but also whether implementing these solutions ultimately sets back or furthers future reforms. Taking this suggestion to its limit, we might adopt an ideal theoretic orientation, first trying to identify an ideally just social arrangement to serve as a “long-term goal of political endeavor,” and then attempting to work out how to make progress toward it.<sup>111</sup> This approach to reform theory, after all, seems tailor-made to accommodating the complexity of our world: to identify a social arrangement as ideal, we must take into account the combinatorial interactions of all its component features, and to determine whether a short-term change constitutes progress toward this ideal, we must take into account all relevant instances of transitional complexity. But, as we will soon see, this approach, too, is unsuitable for a complex world. Whereas an exclusive focus on problem solving is too myopic, we cannot see nearly as far as ideal theory presumes.

### 3.3 The Ideal Target View

In the last chapter we considered and ultimately rejected the ideal benchmark view, on which our conception of the ideally just society serves as an indispensable benchmark for evaluating societies in terms of how far they deviate from it. Along the way, we noted that deviations from the ideal only plausibly track justice if they are deviations with respect to the basic values or principles that are realized in the ideal, rather than deviations with respect to the ideal’s more concrete institutional—or, more generally, social—features. This, we can now see, follows from combinatorial complexity: because different social features interact to

---

<sup>111</sup> Rawls 1999a, p. 138.

produce outcomes, implementing social features that would obtain at the ideally just society does not necessarily make our society more just. But the importance of this point for present purposes is that, as far as ideal benchmark theorists are concerned, the appropriate level of description for the ideal is one that concerns the basic values or principles that explain why it is ideally just. A conception of the ideally just social arrangement may be relevant to the extent that it bears on the plausibility of our basic values and principles. But it is not fundamentally what matters from the perspective of the ideal benchmark view.

When it comes to the ideal target view, however, the situation is reversed. If the ideally just society is to serve as a long-term target for reform, then identifying the ideal in abstract, evaluative terms—as “whichever society maximizes freedom,” or “whichever society best balances efficiency and equality,” or “whichever society best realizes Rawls’s principles of justice,” or so on—will not do. We must also identify the concrete social features that would be realized in the ideally just society, so that we can begin to form judgments about what changes to our current arrangements would help us to achieve the ultimate goal of reaching that ideal social arrangement.<sup>112</sup> This comes out especially clearly when we note that individuals who share criteria of justice may differ entirely in what social arrangements they believe would best realize such criteria. To take an extreme case: if you and I both believe that the ideal society would optimally promote the values of efficiency and equality (and even agree about the tradeoff rate we assign these values) but you believe that a command economy would best satisfy these criteria, whereas I believe that an unregulated market with a minimal state would do the same, we will take ourselves to have very different long-term goals, and so endorse very different social changes as steps in the

---

<sup>112</sup> Compare Gaus 2016, p. 4, who cites Rawls’s 1999b, p. 138 own claim that our conception of the ideally just society must be “importantly institutional” if it is to play this role.

direction of ideal justice. And even setting aside such extreme examples, the point generalizes. We cannot decide which changes bring us toward the ideal unless we have a fairly good idea of what the ideal would look like. Merely aiming, say, to eventually satisfy Rawls's two principles of justice is not enough, since we cannot know which steps bring us toward their satisfaction unless we know which social arrangement would realize them.

So ideal target theorists must aim to arrive at much more concrete descriptions of the ideal than ideal benchmark theorists, given the different functions that the conception of the ideal is intended to play on each view. And this brings us to a second difference between the two ideal theoretic positions: that while ideal benchmark theorists might remain content identifying the most just arrangement that is conceptually possible, this will not do for ideal target theorists. For in order for a conception of the ideal to serve as a long-term goal, it is not enough that it reaches, say, a standard of conceptual perfection, since this standard might be unattainable in the real world. Rather, if we are to treat the ideal as a long-term goal, then it must instead be something that it is possible for us eventually to achieve, in the sense that there is some feasible path between us and it.

Admittedly, this idea is very rough. The feasibility of a social arrangement depends both on what our society is currently like and on which ways of transforming it are compatible with various social scientific (and other) facts; but beyond this, there is ample room for disagreement. Moreover, ideal target theorists typically recognize that it is strictly speaking not enough that we can feasibly achieve the ideal from where we are, since we must also take into account whether any such path to the ideal would involve the violation of a moral requirement. Ideal target theory, in other words, does not claim that the ends always justify the means, but rather “looks for policies and courses of action that are *morally*

*permissible* and politically possible as well as likely to be effective.”<sup>113</sup> And there is again room for disagreement about when a transition from one social arrangement to another is permissible, and therefore about which social arrangements are not only causally but also morally feasible.<sup>114</sup> Still, no matter the details, the point is that our judgment about whether something is feasible depends on a prediction of whether we can get there from here through some morally permissible path.<sup>115</sup> To identify the ideally just society in the sense that is relevant to the ideal target view, we cannot simply form a conception of what a *perfectly* or *fully* just arrangement would be like, since such an arrangement might very well be causally or morally infeasible, and therefore unable to play the role of a long-term goal. Instead, we must form a prediction of which social arrangements we could eventually realize through permissible means, and then another prediction of which, of these, would be most just.

Of course, the goal of ideal target theorists is not merely to identify the ideally just society—understood here, and throughout the rest of this chapter, as the most just social arrangement we could ever achieve through morally permissible means. It is also to identify changes to our current arrangement that would constitute progress toward this ideal. Now, on first glance, it might seem that whereas combinatorial complexity undermines the idea that we make our society more just by realizing more social features in common with the ideal, we can repurpose this idea to provide the following account of “progress toward”: we make progress toward the ideal by making our social features more similar to those realized in the ideal. But, on reflection, this account won’t do. Making our society more similar in its social features to the ideal may not only, due to combinatorial complexity, result in a

---

<sup>113</sup> Rawls 1999b, p. 89, emphasis mine.

<sup>114</sup> Compare Buchanan 2004, pp. 61-63.

<sup>115</sup> Compare Wiens 2015b.

decrease in justice, but may also, due to transitional complexity, set back the achievement of the ideal, since realizing one more social feature in common with the ideal may, for example, trigger a backlash, or lock us into our new arrangement.<sup>116</sup> And this implies that we cannot understand progress toward the ideal in terms of making our current arrangement better approximate or resemble the ideal social arrangement. Instead, ideal theorists must aim to identify “steps” that constitute progress toward the ideal, where such steps may sometimes involve making our arrangement less similar to the ideal along the way.<sup>117</sup>

But how, more specifically, should we understand this idea of “progress toward”? Well, recall Simmon’s earlier worry that implementing solutions to particular problems of injustice might “retard,” “stall,” or “permanently block... movement toward overall justice.” The intelligibility of this worry suggests that “progress toward” should be understood in temporal terms: we make progress toward the ideal by decreasing the time it will take to get there, and progress away by increasing this time, at the limit, making it so that we will never achieve it. But since we rarely know for certain whether a change would make the ideal impossible ever to achieve (more on which shortly), in actual contexts of social reform, we typically must reformulate considerations of possibility in terms of the probability we will ever reach the ideal, so that another way to make progress toward the ideal is to increase the

---

<sup>116</sup> This implication of transitional complexity finds no expression in recent models of the complexity of social reform, such as that found in Gaus 2016 and Page 2018. Though such thinkers are clearly aware of transitional complexity, the models they employ assume that making a social arrangement more similar to another constitutes progress toward the latter. Thanks to transitional complexity, this is not always the case.

<sup>117</sup> Simmons 2010, p. 23.

probability we will eventually get there.<sup>118</sup> This introduces a conflict internal to the notion of “progress toward”: for example, a step down a revolutionary path may have less chance of taking us to the ideal but be faster if it works, while a step down an incremental path might be more of a sure thing yet take longer. But let us put aside such issues, and assume that we have settled on a criterion that aggregates time and probability (and anything else relevant to “progress toward”)<sup>119</sup> into a single standard for judging how “far” the ideal is from a given arrangement. In determining whether a social change would constitute progress to the ideal, we must therefore form a prediction of how the resulting arrangement will continue to change. We must forecast forward from the initial change, asking what effect it will have on the probability we will ever reach the ideal, the time it will take to do so, and the extent to which it thus constitutes progress toward or away from it.

Ideal target theorists must therefore not only identify which social arrangements we could ever (permissibly) achieve and which of these arrangements would be most just, but they must also identify (permissible) steps that bring us toward this ideal in a temporal or probabilistic sense. This, we will see, is a rather herculean endeavor. But before we go on to criticize the approach, it remains to be seen what role pursuing the ideal is meant to play in a full account of social reform. Now, if identifying and pursuing the ideal is at all worth doing, it must be because there are some cases where implementing short-term improvements sets

---

<sup>118</sup> Note here the structural similarity to “conditional probability” models of feasibility, on which we make something more feasible by increasing the probability we will achieve it. See Pablo and Lawford-Smith 2012 and Lawford-Smith 2013.

<sup>119</sup> For example, we might also factor in the morally relevant cost of achieving the ideal. See Rääkkä 1998. But I prefer to analyze such costs separately, as something to be traded off against making progress to the ideal rather than an element of such progress.

back progress toward the ideal. Otherwise, there would be no need to identify an ideal in order to make progress toward it: every short-term improvement would simultaneously constitute progress toward the ideal, so we could make progress toward the ideal simply by identifying and implementing such short-term improvements via problem solving.<sup>120</sup> But, as we have seen, combinatorial and transitional complexity do generate a genuine tradeoff between short-term justice and progress toward the ideal, so there are indeed cases where pursuing the ideal comes at the expense of ameliorating present injustice. This, however, is not to say that we should care only about progress toward the ideal, and we should not saddle ideal theorists with such an extreme commitment. As we have seen, ideal target theorists leave room for the idea that there may be times when ignoring or aggravating short-term injustice in order to make progress toward the ideal is morally impermissible, or simply not worth the cost.<sup>121</sup> So we should understand ideal target theorists as claiming not that we ought always to pursue the ideal at the expense of short-term justice, but that we ought to do so sometimes. They need only claim, in other words, that there is some non-trivial range of cases where the expected long-term benefit of pursuing the ideal outweighs the expected short-term cost of foregoing a short-term improvement, or even of making things worse before they get better.

So understood, ideal target theory is both maximally comprehensive and maximally long-term, and it might therefore seem to fully accommodate the complexity of our world. Working out what social arrangement is ideal requires us to take into account all the ways that social features might combine to produce justice-relevant outcomes, thus fully accommodating combinatorial complexity. And working out steps that constitute progress

---

<sup>120</sup> This is a central message of Gaus 2016.

<sup>121</sup> Rawls 1999b, p. 89, Simmons 2010, pp. 18-22.

toward this ideal requires us to take into account all the ways that earlier changes might impact later ones, thus fully accommodating transitional complexity. But, alas, the very complexity that makes ideal target theory attractive also makes it impracticable—at least for agents anything like us. For, as I will now argue, combinatorial and transitional complexity not only ensure that short-term solutions to particular problems of injustice sometimes conflict with progress toward the ideal, they also give rise to two epistemic asymmetries. First, due to combinatorial complexity, as we consider larger changes to (more features of) our social arrangement, our predictions of what outcome those changes will produce, and therefore of what their effect on justice will be, decrease in reliability. Second, due to transitional complexity, as we forecast the effects of social changes further into the future, our predictions about which further changes will occur become less reliable as well. And, as I will further argue, these two asymmetries undermine the epistemic presumptions of ideal target theory. We cannot identify the ideal social arrangement with sufficient confidence to warrant pursuing it at the expense of short-term justice. And even if we could, we would still lack the epistemic wherewithal to identify changes to our current arrangement that would constitute progress toward it with this requisite degree of confidence.

#### 3.4 Complexity and The Ideal Target View

Consider first a recent argument of Gerald Gaus's, which begins by noting that our (often implicit) predictive models of how social arrangements interact to produce outcomes are not especially accurate in the first place—they not only come with a probabilistic margin of error, but may also fail to assign any probabilities to wholly unanticipated consequences<sup>122</sup>—

---

<sup>122</sup> These two possibilities track Knight's 1921 distinction between probabilistic "risk" and non-probabilistic "uncertainty."

and, furthermore, that like all models of complex systems, they are subject to “error inflation.”<sup>123</sup> We may calibrate our models of our current social arrangement to the data. If we predict, say, that increasing the minimum wage will increase unemployment, or that improving pay for public school teachers will promote better educational outcomes, but find that this doesn’t actually occur—or that either change produces some wholly unanticipated effect, say, on racial inequality (perhaps because, due to an interaction with existing informal norms, raising the minimum wage leads employers to lay off employees of certain social groups but not others)—we may go back and revise our model in light of this feedback. But when it comes to models of merely hypothetical arrangements, we cannot calibrate our models in this way, and so are more prone to error. Error inflation then occurs as we consider social arrangements that differ more and more in their features from actual ones. Gaus explains: “An error in predicting the workings of one feature will spread to errors in predicting the justice-relevant workings of interconnected features, magnifying the original error. As this new erroneous model is used as the basis for understanding yet further arrangements, the magnified errors become part of the new model, which is then itself subject to the same dynamic.”<sup>124</sup>

---

<sup>123</sup> Gaus 2016, p. 80. Smith 1998 provides an enlightening discussion of error inflation, though in the context of chaotic rather than complex systems. The contrast here is that chaotic systems necessarily exhibit extreme sensitivity to initial conditions, but needn’t contain many interacting parts, whereas complex systems necessarily contain many interacting parts and typically contain moderate sensitivity to initial conditions, but only exhibit extreme sensitivity to initial conditions at very high levels of complexity—levels that, I argue in the next chapter, do not typically characterize social systems. See Page 2010, ch. 1.

<sup>124</sup> Gaus 2016, p. 80.

The upshot of error inflation is clear. We should have more confidence in our prediction of the effect of a change, say, to either the minimum wage or pay for public school teachers, than in our prediction about a change to both, since the errors we make in predicting the effect of each carry over into our prediction of what will happen if both changes occur, and how these changes will interact. (To return to our earlier example, perhaps we would fail to predict that raising the minimum wage would aggravate racial disparities in wealth to the point that minority races would become increasingly priced out of wealthier neighborhoods, and that improving pay to public school teachers would lead to understaffing and so to worse rather than better educational outcomes in schools in poorer neighborhoods, such that the two changes together would aggravate racial disparities in education and so, despite our best intentions, undermine equality of opportunity.) We should have more confidence in this than in our prediction of how a radically redesigned economic system such as market socialism or property-owning democracy would work out in practice, since such systems differ from actual ones in so many ways that errors massively inflate. (In addition to increasing the number of the above sort of interactions involving more traditional issues of economic and racial justice, we must also factor in, for example, how such systems would impact technological innovation and the environment, how technological and environmental changes would impact each other, how the benefits and costs of technological and environmental disruption would be distributed, and so on). And we should have even less confidence in our ability to predict the outcome produced by an even more divergent social arrangement, designed to handle not only problems of economic injustice, but also problems of racial and gender injustice, environmental justice, and all other forms of injustice as well. (How would the above sorts of changes interact with further changes to racial or gender norms, to the criminal justice system, to measures designed to

mitigate climate change, to our educational institutions, labor and employment legislation, healthcare system, and so on? And how would all of these changes interact with each other? (The mind boggles.) Indeed, we should have basically no confidence in this at all.

An ideal target theorist, seeing where this argument is going, might object that we can calibrate our models not only to data gathered from our current social arrangements, but also to data gathered about other past or present arrangements that we may obtain via social change and experimentation and through cross-societal and -historical analysis. And this is certainly true, and is something that will be important to keep in mind later. But it is still not enough to undermine Gaus's critique. The problem is that many candidate ideals differ greatly in their features not only from our current social arrangements, but also from any that have ever existed. And this poses a dilemma for ideal theorists. On the first horn, if the ideal turns out to differ greatly from any actual arrangement, then, thanks to error inflation, we cannot make a confident prediction about what outcome it will realize, nor, therefore, about how just it will be. On the second horn, if the ideal happens to be fairly similar to some actual past or present arrangement, such that we can judge with reasonable confidence what outcome it would produce, even then, we cannot judge with any confidence that it is most just without comparing it to other dissimilar arrangements whose justice we have little epistemic access to. So, either way, we cannot confidently identify the ideal, because the ideal is the *most just* social arrangement we could ever get to (through permissible means), and we must compare all feasible arrangements before we can deem one most just.<sup>125</sup> An essential

---

<sup>125</sup> An ideal target theorist might attempt to slip between these horns by holding that there is a logical upper limit to justice. For example, perhaps ideal justice is achieved when we have completely eliminated various forms of injustice such as inequality and oppression. And, in that case, if a relatively similar arrangement to our own turns out to meet this upper limit, we

presupposition of ideal theory therefore cannot be met: we cannot judge with sufficient confidence which social arrangement is most just for this ideal to be worth pursuing at the expense of short-term justice, as ideal target theorists require.

Gaus's epistemic critique of ideal target theory appeals only to combinatorial complexity. But in case one is not yet convinced, we may considerably strengthen it by appealing to transitional complexity as well. As I have emphasized, identifying the ideal requires us not only to evaluate social arrangements in terms of their justice, but also to determine which of these arrangements is causally and morally feasible—something we could eventually get to through permissible means—and so capable of playing the role of a long-term goal. And identifying whether particular short-term changes constitute progress toward the ideal requires us to predict whether implementing them would increase the probability that we will ever realize it or decrease the time it would take to do so. In other words, ideal target theorists must do three things: confidently determine which social arrangements we could eventually get to through permissible means, confidently determine may identify it as ideal without needing to compare it to other societies after all. But, in the first place, such theories of justice are not very common, since most allow that there are at least some factors relevant to justice that contain no logical upper limit: for example, how free or well off people are. And, in the second, it is manifestly implausible to think that that any actual social arrangement is very similar to one that reaches the *logical* upper limit of justice. Perhaps more plausible is the view that some actual social arrangements are close to the *feasible* upper limit of justice—but then we are back to the dilemma outlined in the main text, since, even if the most just feasible arrangement is rather similar to some actual arrangement, we cannot form a well-grounded judgment about this without examining other feasible arrangements that are quite different.

which of these is most just, and confidently determine which short-term changes constitute morally permissible progress toward this ideal. Gaus's argument suggests that the second of these tasks is beyond our epistemic ken. But this does not tell the whole story. Once we take into account considerations of transitional complexity, we realize that we cannot carry out any of these three tasks with sufficient confidence to warrant pursuing our conception of the ideal at the expense of short-term justice—let alone all three.

As a perfectly general matter, our ability to predict the future becomes less reliable as we attempt to forecast further in time. The relevant mechanism here is once again error inflation: the errors we make in predicting what will happen tomorrow get carried over into our prediction of what will happen the day after, which get carried over into our prediction of what will happen next week, next year, next decade, next century, and so on. And this general tendency is magnified in complex systems, where, due to transitional complexity, it becomes impossible for us confidently to predict anything in the very long-term. In the first place, doing so requires us to predict where backlash will occur, where we will get locked in, and, more generally, how people will respond to changes to our social arrangement by producing further changes, and others to those changes, and so on, far into the future. But these predictions are notoriously difficult to make, not only because each depends on our prior prediction, but also because predicting individuals' responses to social changes requires us to predict what outcome those changes will realize in the first place, how coalitions will form in response to them, how demographics will change, and so on. And, in the second, this requires us to predict the occasion and effect of technological innovations and external shocks. Yet it is deeply implausible to think that we could have predicted the occasion of or social change caused by the invention of the printing press, telephone, radio, television, or Internet, or by the occurrence of the industrial revolution, either world war, nor, going

forward, climate change. And it is similarly implausible to think that we can reliably forecast which further technological changes or external shocks will occur. As a result, as we attempt to forecast further into the future, not only do our probabilistic judgments of how individuals will respond to changes multiply together to decrease our confidence in which further changes will occur, but the probability of totally unexpected events increases as well. The upshot is that it is impossible for us to forecast social change far into the future at all: we cannot determine which social arrangements we could get to in the very long-term, nor which short-term changes bring us toward them, at least with any reasonable degree of confidence.<sup>126</sup> And the fact that we must also make sure not to set down morally impermissible paths makes this all the more difficult.

On the basis of similar considerations, David Wiens concludes that “given the number of variables to which our feasibility assessments must be sensitive, the complexity of their interactions, and the potential for path-dependence, determining whether any particular long-range objective is feasible is beyond human cognitive capacity.”<sup>127</sup> In other words, whereas Gaus worries that we cannot figure out how just various candidate ideals would be, Wiens worries that “we simply cannot determine with any confidence whether particular long-range objectives are feasible, let alone with sufficient confidence to justify adopting a political ideal as a reform target.”<sup>128</sup> And since identifying the ideal requires us to figure out both of these things—to determine which *feasible* social arrangement is *most just*—this suggests that ideal target theory is an impossible enterprise. Indeed, the problem is even

---

<sup>126</sup> Compare Hayek 1960, ch. 2 and Tetlock 2006. I discuss Tetlock’s work on the limits of prediction in more detail in the next chapter.

<sup>127</sup> Wiens 2015b, p. 467.

<sup>128</sup> Ibid.

worse than Gaus or Wiens notes, since even if we somehow knew which social arrangement was ideal, we would still be unable confidently to identify which short-term changes would constitute progress toward it. To do so, we would again have to forecast whether such changes would increase the probability of us ever achieving the very long-term goal of reaching the ideal or decrease the time it would take to do so. But, as we have just seen, we cannot confidently forecast the effect of social changes far into the future in this way: all the same reasons we cannot confidently determine which social arrangements are feasible in the very long-term are similarly reasons for thinking that we cannot confidently judge which short-term changes constitute progress toward them. Backlash alone illustrates the problem, since backlash can result in changes that appear to be going in one direction actually causing the reverse, and because predicting backlash in part requires us to predict what outcome will be realized—something we cannot do with any reasonable degree of confidence when it comes to social arrangements that are very dissimilar to our own. And once we factor in considerations of lock-in, technological change, external shocks, and so on, our epistemic situation looks even more hopeless.

The ideal target view, then, requires us not only to identify which social arrangement, of all the arrangements we could ever permissibly realize, would be most just, but also to determine the extent to which shorter-term transitions constitute progress toward this ideal. But given our epistemic circumstances, we simply cannot do this: we can neither identify the ideal nor identify steps that would constitute progress toward it, at least with sufficient confidence to warrant taking these steps at the expense of short-term justice. The ideal target view is therefore epistemically infeasible. Despite its lofty ambitions, it does not provide us with a viable approach to theorizing about long-term social reform.

### 3.5 The Uses and Abuses of Ideal Theory

Ideal theorists typically argue that our conception of the ideal is practically relevant. It provides us either with a benchmark of evaluation, or with a long-term target for reform. Critics of ideal theory reject both of these claims, arguing that ideal theory is useless and perhaps even harmful. So far from serving as an indispensable benchmark for evaluating nonideal societies, a conception of the ideal may distort our theorizing about social reform. And the ideal does not provide us with a viable long-term target. Its pursuit is more likely to make things worse than to make them better.

With respect to the ideal benchmark view, we have seen (in Chapter 2) that there are grounds for adopting a moderate position. Ideal theory has no priority over theorizing about criteria of comparative justice or the application of such criteria in problem solving. But critics go too far if they claim that ideal theory is completely useless, since it may indeed facilitate reform theory by helping us to develop our evaluative criteria, to generate novel solutions, and to appreciate the problems we face. Still, we have also seen that we should be wary of an overreliance on ideal theory. If we give ideal theory a central role in theorizing about social reform, we are liable to end up with distorted criteria of comparative justice, with unrealistic proposals for reform, and with an inadequate understanding of actual problems. So ideal theory is somewhat helpful, but it can only play a peripheral role in problem solving, lest it does indeed become harmful. Theorizing about the ideally just society is roughly as useful as theorizing about dystopia. Neither sort of theorizing provides a substitute for empirically investigating social actualities, nor for theorizing about hypothetical but nonideal societies.

When it comes to ideal target view, we have now found (in this chapter) that the case against ideal theory is stronger. To pursue the ideal as a long-term target, we must be able to

identify the most just social arrangement we could ever permissibly achieve, and must be able to describe it in fairly concrete terms. But, as we have seen, we simply cannot identify this ideal social arrangement, nor track our progress toward it, with much confidence at all. So it would be a grave mistake to pursue this ideal at the expense of the sort of short-term improvements we can identify through problem solving. As Gaus forcefully puts the worry:

[W]hy should we forgo opportunities to create a more just social world so that we can pursue an uncertain ideal? Those who bear the cost of this pursuit will live in a less just world—their pleas must be discounted. The ideal theorist is convinced that we can give meaning to our political lives by pursuing an inherently uncertain ideal, turning our backs on the pursuit of mundane justice... [F]or us to be under the sway of an ideal theory is for us to ignore relatively clear improvements in justice for the sake of a grander vision for the future. And yet this grand vision is ultimately a mirage, for as we move closer to it, we will see that it was not what we thought it was, and in all probability we can now see that a better alternative lies elsewhere.<sup>129</sup>

So pursuing our conception of the ideal may be harmful: it involves taking a bet that setting back short-term justice will help us eventually achieve an ideal that we cannot identify or track our progress toward with enough confidence to warrant making this bet. And so, to the extent that the widespread practice of ideal theory risks encouraging the active pursuit of the ideal—either gradually in the way I have generally been assuming, or, even worse, all at once through violent revolution—it becomes a rather dangerous enterprise indeed.<sup>130</sup>

The above quote from Gaus not only highlights the danger of following the recommendations made by ideal target theorists, but also brings out an additional

---

<sup>129</sup> Gaus 2016, p. 143.

<sup>130</sup> *Ibid.*, pp. 88-89.

consideration that has been lurking beneath the surface of our discussion so far. Pursuing our conception of the ideal is especially foolish, given that this conception is liable to change over time. Writes Karl Popper:

The method of first establishing an ultimate political aim and then beginning to move toward it is futile if we admit that the aim may be considerably changed during the process of its realization. It may at any moment turn out that the steps so far taken actually lead away from the realization of the new aim. And if we change our direction according to the new aim, then we expose ourselves to the same risk again.<sup>131</sup>

What Gaus and Popper have in mind here is the idea that even if we were somehow able to work out how to make progress toward our current conception of the ideal, we would, in the process of making progress toward it, almost certainly change our understanding of which social arrangement is ideal. This is in part because social change brings with it new evidence about how social features interact to produce outcomes that often lead us to revise our understanding of which social arrangement would best realize our criterion or criteria of justice. But it is also because social changes often lead us to revise our fundamental criteria of justice, as when we come to realize that there is something amiss with our current arrangement that the criteria we currently endorse can't explain. And all of this only adds to the case against pursuing our conception of the ideal. Not only are we chasing a target that we cannot identify with any confidence from where we are, but our best guess about what this ideal looks like is constantly changing. To attempt to make progress toward the ideal is therefore to chase a moving target. And it is to do so at an unacceptable cost.

This *moving target objection* becomes even more troubling once we note that there is

---

<sup>131</sup> Popper 2013, p. 174.

little reason to believe that there is a fixed ideal in the first place. The effects of a social arrangement depend, as we have seen, on the combinatorial interactions of its institutional features, not only with other institutional features, but also with background conditions. And it follows, as Douglass North puts it, that “institutions adopted for a particular time, even if optimal at that time, may be far from optimal as the human environment changes over time.”<sup>132</sup> Changes to demographics, resources, the natural environment, technology, and so on, may all affect the justice of institutions. Strict norms of equal division may, for example, be necessary for maintaining peace and order in small communities, but may be detrimental in large market societies where they disincentive productivity and lead to leveling down.<sup>133</sup> Relatively unregulated free market institutions may in turn be an optimal institutional structure in cases where there are few market failures, but more regulations may be needed as the economy grows and the environment changes:

There is not much call for employment regulations if there are few employees, and virtually all have a ready exit into self-employment. When no enterprises are large enough to have market power, there is no need for anti-trust regulation. When land is abundant and practically free, land use and pollution regulations are hardly needed because people are spread out and environmental effects... minimal. When people can appraise the quality of virtually all goods for sale on inspection, and nearly everyone grows what they eat, there is little need for laws regulating the safety of consumer goods.<sup>134</sup>

Likewise, changes to technology, such as occurred during the industrial and information

---

<sup>132</sup> North 2005, p. 22; compare Muldoon 2016, pp. 6, 29 and Rosenberg 2016, pp. 64-70.

<sup>133</sup> Platteau 2000, ch. 5.

<sup>134</sup> Anderson 2017, p. 27.

revolutions, and as they occur at an increasingly rapid rate today, have raised new questions and a need for new institutions that had no place in the past—for example, those regulating the use of CO<sub>2</sub>-emitting technology, digital piracy, “fake news,” and increasingly deadly weapons. And history, too, can matter. American Slavery and Jim Crow on the one hand, and the conquering of North America and subsequent treatment of the folks who were already there (and their ancestors) on the other, may certainly be relevant to the justice of institutional arrangements tasked with discharging compensatory justice in the United States today. And nearly every society on Earth has a similarly speckled past.

Summing up our discussion, then, the problem is not only that we lack the epistemic ability to identify and pursue the ideal with the degree of confidence that would justify doing so at the expense of short-term justice, but that we have no fixed ideal to pursue in the first place. Even if there is an ideal institutional arrangement for every given set of background conditions, such background conditions are unlikely ever to stop changing. And so, there is unlikely to be any fixed ideal social arrangement (which, recall, consists of both institutions and background conditions) at all. Pursuing our current conception of the ideal, as ideal target theorists recommend, is therefore a rather hopeless prospect. Thanks to combinatorial complexity, we cannot confidently identify the ideal social arrangement with any reasonable degree of confidence, and the ideal institutional arrangement may change as our background conditions change with it. Thanks to transitional complexity, we cannot confidently identify what current changes to our current arrangement would constitute progress toward our current conception of the ideal, nor which would bring us toward the conception of the ideal we would hold by the time we got there. It would therefore be the height of folly to favor the blind pursuit of our current conception of what is currently ideal over tangible improvements to the actual. Ideal theory may play some limited role in facilitating problem

solving, but it is useless when it comes to providing us with a long-term goal for reform. It becomes harmful if we fail to recognize these limits, and rely on ideal theory at the expense of other sorts of theorizing. And it becomes dangerous when we treat it as providing us with a long-term goal that is worth pursuing even at the expense of short-term justice.

Critics of ideal theory, however, should not be too quick to rejoice. The failure of the ideal target view is a lamentable failure, not a welcome one. Ideal target theorists, recall, propose ideal theory as a supplement to problem solving—one that it intended to remedy what they take to be its defects. And though we have seen that ideal target theorists are wrong to think that ideal theory can clear up such defects, we have also seen that their critique of problem solving is spot on. Problem solving remains an unsatisfactory approach to reform theory, because there is still the worry that implementing narrow solutions to particular problems of injustice is an unreliable way to promote overall long-term justice. The trick will be to see if we can thread the needle, and avoid both the myopia of problem solving and the impracticability of ideal theory.

## CHAPTER 4 – GETTING BETTER AT GETTING BETTER

### 4.1. A Methodological Challenge

We have now seen that, in theorizing about social reform, we face two competing pressures. Because of the potential for interactions between targeted solutions to particular problems of injustice, there is a pressure to expand our sights outward and forward toward more comprehensive changes and their longer-term effects. But at the same time, our ability to predict the effect of social change rapidly deteriorates as we attempt to expand our sight in both of these ways. So there is a contravening pressure to contract our focus back to the narrow and the short-term.

Now, the problem raised by these competing pressures is not merely one for ideal target theorists, nor is it one that depends on the precise limits of our predictive capacities. Even if we were fairly adept at predicting the effects of and tracking our progress toward, say, medium-sized, medium-term changes, there would remain a gap between these largest and longest-term reforms whose effects we could confidently identify and pursue, and more comprehensive social changes that would produce greater justice in the longer term. And implementing such medium-sized, medium-term changes would still risk setting back the pursuit of overall long-term justice, by interacting either to undermine overall justice (due to combinatorial complexity) or to set back the prospects of future improvements (due to transitional complexity), just like narrow, short-term changes. This leaves us with the following methodological challenge: *how can we identify social reforms that we can predict with reasonable confidence will promote overall long-term justice, when we cannot figure out which social arrangements we should aim at in the long-term, nor which changes to our current arrangements would constitute progress toward them?*

Although problem solvers rarely address this question directly, they sometimes

appear to maintain that we cannot meet this challenge, and that we should therefore scale back our ambitions and carry on with problem solving while taking into account combinatorial and transitional complexity to the limited extent that we can. For example, Wiens writes that when we engage in problem solving, we should “avoid negative interactions as far as possible” and do our best to “keep open possibilities for future improvement.”<sup>135</sup> But since we are not very good at this, we must recognize that the solutions we generate are “tentative and experimental.”<sup>136</sup> We must give up on anything approaching certainty and, as Elizabeth Anderson puts it, treat “imagined solutions to identified problems... as *hypotheses*, to be tested in experience.”<sup>137</sup>

Problem solvers who endorse this experimental orientation are on the right track. We do need to adopt an experimental attitude when theorizing about and pursuing social reform; we do need to think of “reforms as experiments”—to borrow Donald Campbell’s instructive phrase.<sup>138</sup> But, as I will argue in this chapter, a full appreciation of the limits of problem solving mandates a far more significant shift in our overall approach to theorizing about social reform than even experimentally-oriented problem solvers appear to realize. In particular, I will argue that it pushes us to a new approach to theorizing about social reform, which I call the *progressiveness approach*. On this approach, we follow ideal theorists in thinking that we must supplement problem solving with a form of theorizing that focuses on making

---

<sup>135</sup> Wiens 2012, p. 66.

<sup>136</sup> Ibid.

<sup>137</sup> Anderson 2010, p. 6. On Anderson’s view, we should think of *ideals* in this way. But I avoid this terminology to ward off the potential for confusion, since, as always, I reserve the term “ideal” as shorthand for “the ideally just society.”

<sup>138</sup> Campbell 1969.

changes to our society that promote overall long-term justice. But rather than doing so by attempting to identify ways of making *progress toward* the ideally just social arrangement, we instead approach questions of long-term justice by working out how to make our current social arrangement more *progressive*: better at getting better, or more conducive to further improvements in general. And, more specifically, I will argue that the progressiveness of a social arrangement depends on its ability to flexibly experiment with many promising solutions to problems as they arise, to select for or stabilize those solutions that prove successful while eliminating those that do not, and to help us learn from both our successes and our inevitable failures. So while problem solving does have a place on the progressiveness approach, the solutions it generates are viewed as hypotheses that function within a broader framework of experimentation, selection, and learning. Ideal theory, however, has no larger role than on the problem solving approach: theorizing about overall long-term justice instead takes the form of figuring out how to enhance (or at least maintain) the progressiveness of this overarching framework.

I begin by explaining why a focus on progressiveness in general, and on experimentation, selection, and learning in particular, allows us to overcome the difficulties that combinatorial and transitional complexity raise for the problem solving approach, and so to identify reforms that promote long-term justice without recourse to ideal theory. I then render these ideas more concrete by connecting them to a few existing research programs—those concerning experimental democracy, polycentricity, and experimental governance—that, I suggest, can each be understood as an attempt to work out what a reasonably progressive social arrangement looks like in practice. From here, I draw together these threads to explicate the progressiveness approach and defend it from the worry that it runs into the very epistemic difficulties that I have claimed beset ideal target theory. This

completes my argument that the progressiveness approach provides us with a response to our central methodological challenge, and therefore provides a better way to theorize about social reform than either the problem solving approach or ideal theory.

My primary goal in this chapter, then, is to explain and defend the progressiveness approach. So, for the most part, I refrain from coming down on substantive questions about which particular changes to existing societies would make them more progressive. This leaves me open to the worry that while the progressiveness approach sounds good on paper, it isn't very fruitful—that when we attempt to apply it, we will find that it tells us very little about how we should reform actual societies. Although a full response to this challenge must await future applications of the progressiveness approach, I attempt to mitigate this worry at the end of the chapter, where I begin to explore how we might go about working out how to improve the progressiveness of existing societies, either by defending criterion-dependent reforms (that improve progressiveness only as understood by some particular criterion of justice) or, even better, criterion-independent reforms (that improve progressiveness given a wide range of plausible criteria of justice). After that, I will be well-positioned to provide my final argument, both against practically-minded and anti-practical ideal theorists, that political philosophers ought to widely adopt the progressiveness approach in lieu of ideal theory, both so that we can better understand justice, and so that we can better pursue it.

#### 4.2. From Problem Solving to Progressiveness

Just above, I endorsed the suggestion that we adopt an experimental orientation when theorizing about social reform. Let us now consider why such a change is a welcome one, and what such a change involves. Recall from our discussion of the ideal target view that, when engaging in reform theory, we are subject to two epistemic asymmetries: we are worse

at predicting the effects of larger changes than smaller changes, and worse at predicting the longer-term effects of changes than their shorter-term effects. Indeed, at the limit, we are worse at predicting the effect of any change than we are at evaluating our current arrangement, since every social change brings with it some probabilistic margin of error and some risk of totally unanticipated consequences. Thankfully, these asymmetries have a flip side. We are better at evaluating social changes after they are implemented than we are at predicting what they will do, better at evaluating where we have ended up than predicting where we will go. And this explains why the test of any proposed solution's effect on justice must be how it works out in practice. Problem solving can at best serve as a means of hypothesis generation. Its goal must be to discover social reforms that are “worth a try,”<sup>139</sup> but whose actual effects on justice can only be ascertained through trying them out in various combinations—that is, through social experimentation.<sup>140</sup>

Once we recognize this much, however, we cannot stop here and rest content carrying on with problem solving as we described it in the last chapter. To begin to see why, let us consider what makes a proposed solution a good one. What determines, in other words, when a prospective reform is a *promising* one, in the sense that it is genuinely worth a try? One factor is still the effect we predict the reform will have on one or more present problems of injustice: on its amelioration of injustice in the short-term. But that is not all.

---

<sup>139</sup> Schmitz 2015, p. 772. Technically, Schmitz claims that *ideals* must be worth a try, but he uses the term to refer to an ideal solution to a problem rather than to an ideal social arrangement. I avoid this terminology, once again, to prevent confusion.

<sup>140</sup> An emphasis on social experimentation as a means for promoting reform is nothing new, but is a running theme throughout the history of political philosophy. See especially Mill 1977, Dewey 1927, and Popper 2013.

For one thing, it also matters how much experiments reveal about how to improve justice in the future, since some experiments provide more useful information than others that we may feed back into our (explicit or implicit) causal models in order to develop more promising solutions going forward. Most obviously, we learn more from social reforms that produce novel combinations of social features than from those we are already familiar with: since we are better at ex post monitoring and evaluation than at ex ante prediction, implementing such solutions allows us to generate new information by *exploring* the space of social possibilities rather than merely *exploiting* current knowledge of what has worked in the past.<sup>141</sup> Perhaps less obviously, we must also take into account the epistemic quality of the experiments in which we engage. While the gold standard of randomized controlled trials are often unavailable in these contexts, social scientists, program evaluators, and philosophers of science have designed a range of experimental and “quasi-experimental” designs that allow us to draw reasonably strong causal inferences about the effects of social reforms, including those that are specifically designed to take into account complex interactions between variables.<sup>142</sup> This is not the place to go into the technical details of such approaches, but the key takeaway is that, in thinking about which experiments are worth a try, problem solvers must consider not only their predicted effects on justice, but also how informative they are.

---

<sup>141</sup> March 1991 provides a classic discussion of the exploration/exploitation tradeoff, but see also D’Agostino 2009b.

<sup>142</sup> For a start, see Shadish, Cook, and Campbell 2002, who provide a thorough discussion of the relative feasibility and merits of a number of such designs, and Rossi, Lipsey, and Freeman 2003 who provide a helpful overview aimed at practicing program evaluators. See also Cartwright and Hardie 2012, who lay out various techniques we may employ to infer from the fact that a reform worked in one particular context that it will also work in another.

Sometimes the more promising social reform may be the one that would generate more useful information than the one that we are more confident will increase justice in the short-term but without teaching us anything going forward.

This might seem more like mad science than social reform. Why should we trade off the amelioration of present problems of injustice against the extent to which they will enhance our understanding of how different social features interact to produce outcomes? The answer is that we face a tradeoff between short-term and long-term justice, and that the better we understand how social arrangements produce their justice-relevant effects, the more likely we are to promote justice in the future—at least on the assumption that the “we” in question is more apt to use this information for good than for bad (more on which in a moment). Of course, we should not always forgo short-term justice for this reason, maniacally expanding our knowledge about how to improve justice without ever putting it to use. We should sometimes resolve the tradeoff in favor of short-term justice, and we should certainly refrain from experiments that are morally impermissible—say, because they impose severe risks on individuals who do not voluntarily bear them,<sup>143</sup> or because their costs systematically fall on already disadvantaged groups.<sup>144</sup> But just as ideal theorists argue that we must sometimes forego short-term justice in order to make progress toward the ideal, my suggestion is that we must sometimes forego short-term justice in order to better position ourselves to promote further reform. We must trade off the predicted short-term justice produced by a social change not against how much it constitutes *progress toward* the ideal, but against how *progressive* it is: how conducive it is to further improvements in general, though not necessarily to the achievement of any antecedently specified long-term goal. And one

---

<sup>143</sup> Müller 2019, pp. 176-177.

<sup>144</sup> Knight and Johnson 2011, pp. 48-49.

factor that is relevant to the progressiveness of a social arrangement is how much we have learned, and how well we are able to learn, about how to improve justice from there.

This brings us to another way of theorizing about long-term justice that does not qualify as either problem solving or ideal theory. In particular, while improving our understanding of how social features interact through social experimentation is one way for us to improve progressiveness, we can similarly improve progressiveness by enhancing the framework within which experimentation takes place—by making it more amenable to learning. In part, we might do so by improving individual epistemic abilities, for example, through education. But at the institutional level, we might also improve the social epistemic conditions in which we theorize about, discuss, and pursue reform. To some extent, this depends on the existence of conditions that are conducive to good inquiry in general, including the protection of free speech, a diversity of research agendas, shared vocabularies, and so on, that philosophers of science, at least since Thomas Kuhn, have made much progress exploring.<sup>145</sup> But, as I will discuss in more detail shortly, it also depends more specifically on the extent to which we have mechanisms in place for monitoring the social experiments we engage in, for gathering additional information from the experiences of other past and present societies, for dispersing this information widely throughout society, and for storing it in our individual and institutional memories.

Putting this idea more generally: just as we earlier understood making *progress toward* the ideal as increasing the probability we will achieve the ideal or decreasing the time it will take to do so, we may now similarly understand improving the *progressiveness* of a social

---

<sup>145</sup> Kuhn 1962. See also Kitcher 1993, D’Agostino 2009b, and, for a discussion of social epistemic conditions that are conducive to good *moral* inquiry in particular, Buchanan and Powell 2018.

arrangement as increasing the probability and speed with which it will continue to improve in justice indefinitely into the future. But the progressiveness of a social arrangement depends not only on its amenability to learning, since even a society that is able to generate quite a lot of information about how social arrangements produce their justice-relevant effects is unlikely to be conducive to further improvement if this information is never put to use, or, worse, if it is more often channeled in negative, justice-undermining way than in positive, justice-enhancing ways—as it would, to take an extreme but illustrative case, if the new information always ended up in the hands of a powerful group hell-bent on using that information to line their own pockets at the expense of justice. So progressiveness also depends on the extent to which our arrangement facilitates flexible experimentation with a wide range of promising (though not impermissible or otherwise unpromising) solutions to problems of injustice going forward. Among other things, this may require the replacement of norms of dogmatism with those permitting experimentation,<sup>146</sup> the ability to avoid lock-in due to seizure by interest groups,<sup>147</sup> and a general reluctance to implement reforms that are difficult to reverse.<sup>148</sup> This, too, is a topic to which I will return shortly.

At the same time, since our goal is to promote long-term justice, it would be too simplistic to think that we really always ought to avoid lock-in, that we want to leave all options open going forward. Sometimes, we do want to close options off, at least temporarily: we want mechanisms in place to ensure (or at least increase the odds) that reforms which genuinely solve problems of injustice remain stable for as long as they remain solutions. Indeed, as I have just suggested, it would be naïve to assume that those with

---

<sup>146</sup> This is a running theme of both Dewey 1927 and Popper 2013.

<sup>147</sup> North 2005, p. 125.

<sup>148</sup> Campbell 1969, p. 410.

influence over the political process will always use this influence in order to attempt to promote justice: the entrenchment of successful reforms is one important bulwark against those who would otherwise reverse successful reforms for their personal enrichment. And this mandates a reliance not only on learning mechanisms that help us to learn from our experiments, but also on selection mechanisms by which our society is led to stabilize successful experiments and eliminate failed ones.

To unpack this idea further, let us say that a social feature is “worth keeping” if, taking into account all the practical and epistemic benefits and costs it provides, there is no feasible change to it that is promising, or “worth a try.” And let us say that a social experiment is a “success” if it is worth keeping, a “failure” if it is not. Now, the reason we have to engage in social experimentation in the first place is that we cannot confidently predict which social changes will be successes and failures. And this difficulty is only magnified by the fact that, due to combinatorial complexity, a feature that is worth keeping at one time may cease to be worth keeping at another, after other changes have occurred that interact with it. So we often cannot predict which of the social features we implement will prove worth keeping, nor how long they will remain that way—especially as we attempt to forecast further into the future. What we need, then, are selection mechanisms that allow us to stabilize social features that prove worth keeping, and to modify those that are not. Progressiveness, in other words, depends not only on learning mechanisms of epistemic feedback, but on selection mechanisms of practical feedback as well.

Engaging in a wide range of social experimentation against the backdrop of a social framework that embodies mechanisms of epistemic and practical feedback—of learning and selection—is the only realistic way for agents like us to promote long-term justice in light of the practical and epistemic difficulties that complexity poses. Due to combinatorial

complexity and the epistemic asymmetry to which it gives rise, we cannot assume that simply implementing targeted solutions to particular problems of injustice will lead us to more just combinations of social features in the long-run, but nor can we hope to predict the effects of large-scale long-term changes (or the cumulative effects of many small changes). So the only realistic way for us to determine the outcomes produced by combinations of features, and to achieve such outcomes when they prove to be more just, is to try out various combinations and monitor them after the fact. Due to transitional complexity and the epistemic asymmetry it generates, we cannot hope to predict what the effect of short-term changes will be on the social arrangement we will end up with in the long run. So the only viable way for us to pursue long-term reform is through mechanisms that allow for continual adjustment of our social arrangement on the basis of the feedback we gather from this experimentation. Thus, experimentation and learning mechanisms of epistemic feedback help us to *tame* the difficulties raised by combinatorial complexity by permitting us to try out various combinations of social features, to learn from such experiments, and to reduce the error that goes into our predictions going forward. But the goal of selection mechanisms of practical feedback is not to tame but to *harness* transitional complexity by reducing our reliance on long-term prediction through our ability to correct for errors as we go along. Though we cannot predict where phenomena like backlash and lock-in will occur, the hope is that we can institutionalize mechanisms that correlate backlash (or a functional equivalent) with social features that prove worth changing, and lock-in (or a functional equivalent) with those that prove not to be. That is what selection mechanisms are meant to do.

The progressiveness of our social arrangement would seem, then, to depend on its ability to facilitate experimentation with a wide range of promising reforms, to help us learn from such experiments, and to select for or stabilize successful experiments. And so, when

engaging in theorizing about social reform, we cannot focus only on attempting to solve particular problems of injustice—even while adopting a tentative, experimental attitude toward such reforms—as problem solvers suggest. For, in the first place, the long-term effects of any given solution depend to a significant extent on the progressiveness of the overarching framework in which it is implemented. How much information does the solution generate? How is this information transmitted throughout society? How stable will the reform be if it is a success? How likely is it to reverse if it is a failure? And how will it change if it destabilizes? Answers to these sorts of questions crucially bear on the long-term justice of any given reform, but they depend on the broader mechanisms of experimentation, learning, and selection that we have in place. And, in the second, when taking into account the effects of any given reform, we must also consider its effects on this broader framework's progressiveness. But we cannot do this without thinking hard about what determines the progressiveness of our existing arrangement, and, more specifically, about what would make it more or less conducive to experimentation, learning, and selection.

We are now ready to articulate a parallel argument against the sufficiency of problem solving as that more standardly articulated by ideal target theorists. According to such ideal theorists, we cannot responsibly engage in reform theory by solely engaging in problem solving, because we must also take into account the extent to which any given solution constitutes *progress toward* (or *away from*) the ideal, and we cannot do this without working out a conception of the ideal. I have argued (in Chapter 3) that ideal target theory is a non-starter, since we can't figure out how to make progress toward or away from the ideal social arrangement with the degree of confidence that it requires. But the parallel argument goes like this: we cannot responsibly theorize about social reform by solely engaging in problem solving, because we must also take into account the extent to which any given solution sets

back or furthers the overall progressiveness of the society in which that solution is implemented, and we cannot do this without theorizing about progressiveness. And thus arises a need for a new type of theorizing, concerned not with working out promising “reforms as experiments,” but with figuring out how to enhance the progressiveness of the overarching framework in which such experimentation takes place.

#### 4.3. Progressiveness in Practice

My discussion of selection mechanisms in particular, and of progressiveness more generally, has been rather abstract. To make this all more concrete, let us now consider a stylized contrast between three current research programs, each of which we may here understand as providing a different approach to realizing the sort of mechanisms we have just discussed. The first is experimental democracy.<sup>149</sup> Here, the rough idea is that experimentation is achieved primarily through democratically authorized institutional experiments, learning occurs as individuals with knowledge of these reforms’ effects deliberate and share information with one another about them, and selection occurs primarily through political argument about and voting on which experiments to maintain and which to reverse. Progressiveness is therefore achieved via a sort of centralized democratic experimenter, as opposed to the more canonical central planner on the one hand, or a decentralized mechanism on the other. And such a system is able to improve its own progressiveness over time as it applies this method reflexively—to the very features that provide for experimentation, selection, and learning.<sup>150</sup> Indeed, according to Knight and Johnson, it is

---

<sup>149</sup> For historical defenses of experimental democracy, see Dewey 1927 and Popper 2013.

For more recent defenses, see Anderson 2006 and Knight and Johnson 2011.

<sup>150</sup> Knight and Johnson 2011, ch. 6.

precisely this reflexive capacity that makes democracy so progressive (though they do not use the term): democracy is, they argue, the only institutional form that is able “(1) to coordinate effective institutional experimentation, (2) to monitor and assess effective institutional performance for the range of institutions available in any society, and most importantly, (3) to monitor and assess its own ongoing performance.”<sup>151</sup>

This brings us to the second major approach: polycentricity.<sup>152</sup> This time, the rough idea is that experimentation is achieved not only through consecutive or diachronic experimentation, but also through a number of social features being tried out simultaneously in different jurisdictions that each serves as an independent center of decision-making (thus, the name “polycentricity”). So, for example, in a federalist system, there are a number of distinct political jurisdictions that, though bound together by common federal laws, each have decision-making authority over a range of issues within its territory. Or, at the informal level, different social groups, though bound together by common laws or norms, may simultaneously try out different informal norms over a range of issues where their shared institutions are silent.<sup>153</sup> In each case, experimentation is achieved through different groups employing different combinations of laws, policies, norms, and background conditions at the same time, and especially through the proliferation of new reforms through the efforts of “public entrepreneurs”; selection occurs as groups compete with one another for adherents,

---

<sup>151</sup> Knight and Johnson 2011, p. 169.

<sup>152</sup> Here, we find an approach primarily advocated for by political economists such as Ostrom 2005 and Paul D. Aligica and Vlad Tarko 2011, but also defended more recently by philosopher such as Müller 2019 and Gaus 2016, ch. 4. Aligica 2018 provides perhaps the most important recent statement of this approach.

<sup>153</sup> Gaus 2016, ch. 4.

say, through “foot-voting” in a federalist system, or through attracting members to one’s social group at the informal level; and learning arises through individuals and groups observing the results not only of their own (formal or informal) institutions, but also those of others, so that they can adjust their own institutions accordingly.

A third and final approach arises out of a recent literature on “experimentalist governance.” This program is focused primarily on the reform and structure of administrative agencies in multinational organizations such as the European Union, but the themes are familiar.<sup>154</sup> As Sabel and Zeitlin aptly summarize this style of governance, there are four important components:

First, broad framework goals and metrics for gauging their achievement are provisionally established by some combination of ‘central’ and ‘local’ units, in consultation with the relevant civil society stakeholders... Second, local units are given broad discretion to pursue these goals in their own way... But, third, as a condition of this autonomy, these units must report regularly on their performance and participate in a peer review.... Fourth and finally, the goals, metrics, and decision-making procedures themselves are periodically revised by a widening circle of actors in response to the problems and possibilities revealed by the review process, and the cycle repeats.<sup>155</sup>

Thus, in jurisdictions realizing experimentalist governance institutions, experimentation again occurs both diachronically and concurrently, as different local administrative units try

---

<sup>154</sup> See Sabel and Zeitlin 2012 for a helpful overview, and Sabel and Zeitlin 2010 for a collection of essays discussing various uses of experimentalist governance in the European Union. See also Wolfe 2018 for examples of this sort of governance in the U.S. and Canada.

<sup>155</sup> Sabel and Zeitlin 2012, pp. 177-178.

out a variety of ways of pursuing broad framework goals. Learning occurs through formal monitoring procedures and processes such as peer review, as well as through periodic revisions of the measurement standards used to monitor the achievement of various goals. And selection occurs as local units revise their attempts to achieve these broad goals in light of evidence they have achieved about their own performance and the performance of others, and as such broad goals are themselves revised across time.

We need not settle here the debate between proponents of experimental democracy, polycentricity, and experimental governance. Indeed, it is perhaps not even right to describe the three approaches as in debate with one another at all, since the conflict between them is much less stark than my stylized discussion has just made it seem. Experimental democrats, for example, generally recognize the importance of some degree of polycentricity, as well as the need for not only formal democratic procedures, but also for epistemic and practical feedback mechanisms provided by a “vigilant press, petitions to government, and public commentary on proposed administrative regulations” as well as by social movements engaging in “disruptive demonstrations and legal action.”<sup>156</sup> Polycentrists universally recognize the importance of an overarching (typically democratic) governance structure to oversee the experiments that take place at its various centers of decision-making, facilitating information sharing, minimizing negative externalities, and maintaining competition through jurisdictions by keeping exit and entry costs low.<sup>157</sup> And whereas proponents of experimental governance focus primarily on administrative decision-making, they similarly embrace both polycentricity for local administrative units and allow that broad framework goals, measurement standards, and decision procedures must be continually updated in something

---

<sup>156</sup> Anderson 2010, p. 99.

<sup>157</sup> See especially Müller 2019 for a defense of “polycentric democracy.”

like a democratic fashion.<sup>158</sup>

Thus, whereas experimental democracy emphasizes voting and talking, polycentricity emphasizes competition between jurisdictions or groups, and experimental governance emphasizes bureaucratic decision-making and formal monitoring, these approaches are not necessarily in conflict. There is, after all, no reason for us to make an exclusive choice between learning mechanisms: we can learn through informal discussion and formal monitoring, through public expressions of discontent via social movement activity and through peer review, through drawing inferences about the effects of social features on people's lives through their voting activity and decisions to enter or leave jurisdictions. And one general lesson of complex systems research is that we similarly need not choose between selection mechanisms: complex systems tend to benefit from a variety of selection mechanisms operating at different time scales, as would occur if some selection occurs through political argument and voting, some through competition between jurisdictions, some through administrative discretion, and some through social movement activity.<sup>159</sup> Theorists of progressiveness can therefore learn much from each of these approaches to realizing mechanisms of experimentation, learning, and selection. In coming up with ways of improving their societies' progressiveness, they can pick and choose from elements of these different approaches, or can come up with new ways to synthesize them, as they attempt to

---

<sup>158</sup> Indeed, Zeitlin and Sabel 2010, p. 18, argue that, so far from standing in opposition to democracy, experimentalist governance “may have broader and potentially democratizing effects on the new system of governance taken as a whole”—in particular, that the implementation of experimentalist governance often “destabilizes entrenched forms of authority... in ways that may clear the way for an eventual reconstruction of democracy.”

<sup>159</sup> See especially Axelrod and Cohen 2001, p. 135

apply the insights of these research programs to questions of how to select for the performance criterion of justice in particular.

In any event, my goal in raising these approaches has been only to give some sense of what a reasonable level of progressiveness would look like in practice. A progressive society will need to be one that engages in a wide range of experimentation, both diachronically and simultaneously. It will need to have formal and informal learning mechanisms in place to monitor such experiments, so that it can gather information about their effects. And it will need to have some combination of top-down and bottom-up selection mechanisms in place by which it is able to change in response to this information, eliminating failed experiments and sustaining successful ones. The main takeaway here, then, is that under conditions where we cannot identify an ideally just society to pursue as a long-term goal, our ability to pursue long-term justice depends on our ability to implement these sort of progressive mechanisms of experimentation, learning, and selection. And since we humans find ourselves under such conditions, and are likely to do so for the foreseeable future, our theorizing about long-term justice therefore must involve working out ways of improving the efficacy of such mechanisms.

#### 4.4. Threading the Needle

Drawing all these threads together, there are, on the approach I have outlined, two basic tasks for theorists of social reform. The first is to engage in problem solving: attempting to identify, to the best of our ability, social changes that would ameliorate particular instances of injustice. But, given the complexity of our world and the epistemic limitations it generates, we must recognize that we cannot really come up with surefire solutions, so much as hypotheses that are worth a try. And in evaluating such hypotheses, we must take into

account not only the extent to which we predict they will solve such problems, but also the extent to which they affect our progressiveness or prospects for future reform going forward—for example, how much we will learn from them, and how difficult they will be to reverse. Or, more to the point, if we live in a democracy whose mechanisms of experimentation, learning, and selection depend, say, on public trust and participation in the democratic process, then we must count it against a divisive policy that passing it would lead to public distrust or alienation from the process—even if the policy would genuinely ameliorate injustice in the short term, its divisiveness notwithstanding.

The second basic task is to theorize directly about how to improve the progressiveness of our social arrangement: the speed and reliability with which it will continue to improve in justice indefinitely into the future. This, I have argued, depends largely on its conduciveness to a wide range of promising experiments, on its ability to select for successful ones while eliminating failed ones, and on its facilitation of learning from both successes and failures. So while ideal theorists are right that problem solving is not enough, they are wrong that we need to supplement problem solving with ideal theory. Instead, we must supplement problem solving with theorizing about how to make our society more progressive. And we must trade off improvements in short-term justice not against progress toward the ideal, but against progressiveness more generally.

That there are only two basic tasks on the progressiveness approach remains, however, compatible with there being other subsidiary tasks that inform these basic ones. For example, one such task may be pure normative theorizing about which abstract criteria of comparative justice to use when evaluating outcomes as more or less just: unless one accepts a version of methodological particularism (discussed in Chapter 2), such theorizing will often be relevant to problem solving, and sometimes also to theorizing about

progressiveness (more on which in the next section).<sup>160</sup> So the progressiveness approach does not rule out other types of even very abstract sorts of theorizing about justice, but is rather distinguished by the two basic tasks with which it is concerned. It differs from the problem solving approach not because it sees problem solving as unimportant, but because it sees problem solving as only one of two basic tasks. And it differs from the ideal theoretic approach because it gives ideal theory no central role, though it recognizes the importance of ideal theory insofar as it facilitates either of its more basic tasks by helping us to develop criteria of comparative justice, to notice problems in our current society, or to come up with novel reforms. The progressiveness approach is therefore more expansive than the problem solving approach, and it deemphasizes without wholly doing away with the focus of the ideal theoretic approach. This allows it to accommodate a concern with both short-term and long-term justice, and in a way that remains within the limits of our epistemic capabilities.

Still, one might worry that the progressiveness approach has only spurious appeal. I have argued that progressiveness depends on experimentation, learning, and selection, and that, given our inability to identify an ideal target, working out how to improve progressiveness through implementing or enhancing such mechanisms is therefore the best way of working out how to promote long-term justice. But can we really do this? Do our epistemic capabilities permit it? Or do the reasons we have to think that ideal target theory is epistemically infeasible similarly count against the progressiveness approach? This is the epistemic objection to the progressiveness approach, and it comes in both a more general and a more specific flavor.

---

<sup>160</sup> For illustrations of the sort of abstract, pure normative theorizing I have in mind, see Barrett 2020b, 2020c.

*The General Epistemic Objection*

The general epistemic objection calls into question the feasibility of both basic tasks recommended by the progressiveness approach: problem solving and theorizing about progressiveness. In particular, we saw in the last chapter that considerations of complexity undermine our ability to work out how to make progress toward the ideal. But if things are so complex, then it might seem that we have little hope for theorizing about social reform more generally. We simply have no idea what changes to our society will do, and so no idea whether they will improve justice or progressiveness. Or so the objection goes.

This objection would be decisive if societies were so complex that they were chaotic or random.<sup>161</sup> In that case, the interactions between social features would be so dense that minor tweaks to them would reverberate throughout the entire arrangement in entirely unpredictable ways (“the butterfly effect”), and we would have no chance of coming up with genuine social reforms at all. But it is important to note that my arguments against ideal target theory in the last chapter in no way committed me to this picture. Instead, my argument there relied only on the claims (i) that as we consider larger changes to our social arrangements and their longer-term effects, our ability to make accurate predictions rapidly declines, due to error inflation (which is a general phenomenon in complex systems), and (ii) that our society is sufficiently complex that, given such error inflation, we cannot confidently identify the ideal (since this requires us to make predictions about massive changes) nor how to make progress toward it (since this requires us to make predictions about very long-term effects). And it remains entirely compatible with these two claims that, (iii), complex systems

---

<sup>161</sup> Theorists of complexity disagree about whether we should more felicitously categorize extremely complex systems as “chaotic” or as “random.” See Page 2010, ch. 1 for helpful discussion. But there is no need to go into such debates here.

are nevertheless not *too* complex to permit us to make relatively reliable predictions about the short-term effects of small changes. Indeed, I have assumed throughout this essay that it is a worthy endeavor to attempt to make exactly such predictions in the context of attempting to come up with reforms that solve or mitigate particular problems of injustice. My argument against problem solving was not, after all, that it is epistemically infeasible to work out narrowly targeted solutions to problems of injustice, but rather that focusing only on such narrow solutions is not a reliable way to promote long-term justice.

So whether or not complexity undermines our ability to engage not only in ideal target theory, but in theorizing about social reform more generally, depends on whether our world is *highly complex*, such that we can't even make small, short-term predictions, or only *moderately complex*, such that the real problem is with making large, long-term predictions. More precisely, it is common to describe complex systems in terms of the number of features  $N$  they contain and the average number of interactions  $K$  between those features.<sup>162</sup> In a maximally complex system, every feature is connected with every other:  $K=N-1$ . Such systems are chaotic, or random, in the sense that there is no correlation between social features and the outcomes they produce: tweaking one small social feature may do just about anything to the outcome achieved. Obviously, in these systems, there isn't much hope for intentional social reform or for learning. The general epistemic objection goes through, and we may as well give up on theorizing about social reform altogether.

As  $K$  decreases, however, and we enter the realm of moderate complexity, things change. In these systems, there are correlations between social features and outcomes, and

---

<sup>162</sup> The analysis that follows draws heavily on Kauffman 1993, Part I, who first developed such “NK” models of complex systems, as well as Gaus 2018, ch. 2, who helpfully applies them to the issues at hand.

small changes to social features tend to have small effects on outcomes.<sup>163</sup> More specifically, features tend to be “nearly decomposable,” in the sense that “the short-run behavior of each [feature] is approximately independent of the short-run behavior of the other components.”<sup>164</sup> So, in moderately complex systems, we can make relatively reliable short-term predictions about the short-term effects of small changes, and learning is possible since experimentation allows us to uncover “schemata” or “building blocks” (combinations of features whose effects on outcomes are nearly independent of other social features)<sup>165</sup> and so “causal principles” (that allow us to describe the conditions under which certain

---

<sup>163</sup> Kauffman 1993, p. 180. As Kauffman explains, exceptions to this general tendency occur when a small change leads to a bifurcation, as when increasing the velocity of a steam flowing past a rock makes little difference to the fluid dynamics until a certain critical velocity is reached, when eddies suddenly begin to form. Such bifurcations are more colloquially known as “tipping points”; see Gladwell 2002.

<sup>164</sup> Simon 1962, p. 474.

<sup>165</sup> Holland 1992, ch. 4. Holland’s insight is that when we observe the effects of a particular combinations of features, we are simultaneously gathering evidence about the effects of a whole range of schemata—all of those that are consistent with that combination. To take a simplified social example: if we experiment with a new tax policy in a rich jurisdiction with low economic inequality, we are similarly gathering evidence about what the policy may do on its own, what it may do in rich jurisdictions, what it may do in jurisdictions that have low economic inequality, and what it may do in jurisdictions that are both rich and characterized by low economic inequality. In this way, we can explore social space in “intrinsically parallel” fashion, and so can learn much faster than we might have otherwise supposed.

combinations of features produce certain effects).<sup>166</sup> After all, if societies are nearly decomposable, if the short-run behavior of social features is approximately independent, then, in the typical case, small changes to individual features only interact in fairly mild ways with other features in the short-term, and predicting such short-term effects should not come with too large a margin of error or too large a risk of something totally unexpected—even though they always come with some such margin and some such risk. It is only when we consider larger changes and their longer-term effects that interactions really start to become epistemically unmanageable as the predictive margin of error increases and the risk of something totally unexpected skyrockets.

In a moderate but not highly complex world, then, theorizing about social reform more generally is possible, even though ideal target theory in particular is not. So, in the present context, the question becomes whether it is more plausible to think that we live in a highly or moderately complex world. And there are a number of reasons to endorse the latter view. The first reason is that, as Herbert Simon has influentially argued, the relative stability of existing social systems and their capacity to evolve across time suggests that they

---

<sup>166</sup> Cartwright and Hardie 2012. To adapt one of their examples, suppose we are concerned with equality of opportunity and so set out to improve education in inner-city schools. We know from past experiments in Tennessee that reducing class size sometimes improves student achievement, but when we try this in California no similar gains occur (p. 4). Why? Because we discover, after further experimentation, that the relevant causal principle says that reduced class sizes only produce better outcomes given sufficient classroom space and qualified teachers, and the latter two factors were absent in inner-city California (pp. 64-65). Of course, actual causal principles are often more complicated, involving more than three factors, but in moderately complex systems they are not always too complicated to learn.

must be not highly but rather moderately complex and so nearly decomposable.<sup>167</sup> After all, as Allen Buchanan and Russell Powell have recently put a similar point, if things were highly complex, a society would be a “house of cards” in which any small change would threaten peace, order, and the proper functioning of our institutions. But societies undergo changes all the time without collapsing or radically altering, so societies simply cannot be this complex. Instead, social features must be nearly decomposable or “quasi-independent” of one another, which, we have seen, is a characteristic of moderately complex systems.<sup>168</sup>

This reveals an inherent tension in the broadly “Hayekian” view according to which societies tend to evolve in beneficial ways across time, but since complexity undermines our epistemic capacity to understand how our current social features produce their effects or to work out ways of intentionally improving upon them, it is generally better to leave things to evolution than to attempt to intervene and intentionally promote greater justice.<sup>169</sup> As a general matter, systems can only evolve if they are moderately complex or nearly decomposable since, as Buchanan and Powell point out—drawing on the evolutionary theorist Richard Lewontin, but echoing Simon—“to be shaped by selection, traits must be capable of modification without disrupting other important components of the system.”<sup>170</sup> So evolution and moderate complexity go together, and with this comes at least some limited

---

<sup>167</sup> Simon 1962, p. 475, *et passim*.

<sup>168</sup> Buchanan and Powell 2018, pp. 263-264.

<sup>169</sup> See especially Hayek 2007. I put “Hayekian” in scare quotes, however, because Hayek’s own views on this issue were rather more nuanced; see Schaefer 2019 for an illuminating discussion. But I have heard the position in the text trotted out as a “Hayekian” challenge to my view enough times that I thought it nevertheless worth addressing on those terms.

<sup>170</sup> Buchanan and Powell 2018, p. 264; Lewontin 1978; Simon 1962.

ability to predict the effect of social change in the narrow and short-term. If one wishes to deny even this predictive capacity, then one must claim that we live in a highly and not merely moderately complex society. But then one must give up on the view that societies are capable of evolution, and so on the above-mentioned “Hayekian” position as well.<sup>171</sup>

Moving on from considerations of stability and evolution, the next major reason to think that we live in a moderately complex world in which theorizing about social reform is not entirely hopeless is provided by history—both by the history of successful social reform, and by the fact that our understanding of how social features interact to produce outcomes has clearly improved across time. We would not have so many examples of successful reforms or of learning if our world were too complex to permit intentional social reform or learning about how social features interact. Perhaps some would retort that successful social reforms have all been flukes, and that we are deluded if we think we really know more about institutional functioning than we used to. I find this hard to believe: just think of your favorite historical social movement—say, the movement to abolish slavery, the civil rights movement, or the women’s liberation movement—or of the various examples I used to illustrate combinatorial complexity back in Chapter 2, which all involved examples of interactions between social features that we now understand much better than we used to.

---

<sup>171</sup> To be clear, near decomposability does not imply that prediction is easy. Indeed, as Gaus 2019, p. 980, points out, nearly decomposable systems may still contain highly complex subsystems. But, again, there is a limit to how complex such subsystems can plausibly be, if they are able to remain stable and capable of evolution. My point here is not that prediction is straightforward or highly reliable in a nearly decomposable system. It is just that things are not so complex as to make attempts to predict the short-term effects of small changes an entirely hopeless endeavor, not even worth engaging in.

For that matter, one might look at resources such as the Campbell Collaboration, which carefully summarize evidence we have about the effects of various reforms in subject areas such as Crime & Justice, Disability, Education, International Development, Knowledge Translation & Implementation, and Social Welfare.<sup>172</sup> But I will not linger on these points, since it would take a much more thorough look at the historical record than it is worth going into here to attempt to convince someone who is inclined to deny them, and they should seem fairly commonplace to all but the most committed skeptic.

In case one does remain so committed, however, a third and final reason to think that our world is moderately but not highly complex involves the empirical evidence concerning our predictive capacities. Take, for instance, the remarkable work of Philip Tetlock, who has perhaps done more than anyone to study the ability of individuals to make predictions about complex political phenomena. Although Tetlock is famous for showing in earlier work that many political experts do no better than random chance at predicting the occurrence of political events, in more recent work he has discovered, through the use of massive prediction tournaments, that certain individuals whom he calls “superforecasters” are astonishingly reliable at forecasting political events given nearly any plausible

<sup>172</sup> See: [www.campbellcollaboration.org](http://www.campbellcollaboration.org). Admittedly, Cartwright and Hardie 2012 criticize the Campbell Corporation for presenting information in a way that suggests that just because we have evidence that a reform works in one context, we can infer that it works in another. In a complex world, this is often not so, since the effect of any given reform depends upon the other social features in place. But, *pace* Gaus 2019, p. 983, the conclusion Cartwright and Hardie draw is not that we can’t learn anything from such evidence; it is that we need to be careful when inferring from the fact that a reform worked “here” to the fact that it will work “there.” And they propose a number of techniques for drawing precisely such inferences.

measurement standard.<sup>173</sup> Furthermore, groups of ordinary forecasters can generate group predictions that approximate the results of superforecasters if we take the average group prediction, or, even better, if we apply certain algorithms designed to correct for certain biases, such as one that takes a weighted average of all individual predictions (with higher weight assigned to the predictions of individuals with better individual prediction records), and then “extremizes” the group prediction by pushing its confidence level further away from 50% toward either 0% or 100%.<sup>174</sup> This level of predictive capacity is inconsistent with a picture of the world on which everything is so complex that we simply can’t predict anything: the evidence shows that certain individuals, and groups of individuals, can indeed predict things rather well. But, in case one begins to worry that this undermines my earlier arguments about the epistemic intractability of ideal theory, it is worth noting that one of Tetlock’s other findings is that even superforecasters do no better than random chance when attempting to forecast more than five years into the future—something we would surely need to do if we wanted to work out how to make progress toward the ideal.<sup>175</sup> So my position is supported not only by complexity theory and by the history of successful social reform and successful learning, but also by the empirical evidence concerning the limits of our predictive capacities as well.

There appear, then, to be a number of principled grounds for holding that our social system is only moderately complex rather than downright chaotic or random. And if this is right, then my approach seems well-suited to our world. Given moderate complexity, we should adopt an attitude of fallibilism, and should trust our predictions about larger and

---

<sup>173</sup> Tetlock 2006; Tetlock and Gardner 2015.

<sup>174</sup> Tetlock and Gardner 2015, pp. 70-71.

<sup>175</sup> *Ibid.*, p. 183.

longer-term changes less than our predictions about smaller and shorter-term changes. But we need not give up on theorizing about social reform altogether.

Before we conclude our discussion of the general epistemic objection, let me briefly consider one further worry: that even if things are only moderately complex now, this is only a temporary state of affairs. After all, it is commonly held that as complex systems evolve across time, they tend to become more complex.<sup>176</sup> So even if arguments I have provided concerning the stability of social systems, our history, and our predictive capacities show that our world is currently moderately complex, they don't show that it won't become too complex for us to work out social reforms at some point in the future—perhaps quite soon, given the rapid rate of technological innovation that characterizes modern society.<sup>177</sup>

Fully addressing this worry would require me to delve quite deeply into various debates in complexity and evolutionary theory, so let me instead simply mention three mitigating factors. First, the mechanisms that lead to increased complexity tend to be reversible. In particular, in his seminal analysis, W. Brian Arthur argues that we should expect not only for complexity to grow across time—for example, as more and more details are added to a tax code, or to a common law system of jurisprudence, as various new issues arise—but also for it to occasionally collapse—for example, when legislation is introduced that greatly simplifies our tax code or a particular area of the criminal law.<sup>178</sup> Second, as we

---

<sup>176</sup> Arthur 2015, p. 144 calls this a “folk theorem, almost” among complexity researchers.

<sup>177</sup> Compare Gaus 2019, pp. 980-981.

<sup>178</sup> Arthur 2015, ch. 9. The only mechanism Arthur identifies that lacks a tendency to reverse is what he calls “capturing software,” which “is the taking over and ‘tasking’ of simpler elements by an outside system for its own... purposes” (p. 152). For example, humans have “learned to ‘capture’ and ‘program’ electrons for our own use,” creating increasingly

continue to explore the space of social possibilities and learn more about how social features interact, our predictive capacities may increase, and this may partially offset increases in complexity. Even if the world becomes harder to predict, we may also become better at predicting. And third, and perhaps most significantly, there are things we can intentionally do to induce near decomposability (and so reduce complexity) in complex systems, through a process called “modularization.”<sup>179</sup> For example, we might introduce certain measures designed to reduce interactions between jurisdictions or that channel them in relatively controlled ways, as when we regulate the particular means by which individuals can migrate and interact (say, through trade) across jurisdictional boundaries. Political constitutions that regiment the procedures by which change can occur and place limits on these procedures (say, by protecting certain constitutionally enshrined rights) may also play a similar role: if, for example, the constitution effectively prevents any political body from enacting policies of a certain type, this can serve to reduce our epistemic burden by taking certain possibilities off the table and allowing us to focus our efforts on smaller regions of social space.

Thus, I believe the real upshot of this last worry is not so much that we will inevitably reach a point where things become too complex to predict, but rather that the complex artifacts powered by electromagnetism (p. 153). Or, to take a more germane example, the financial community has learned to capture and program simpler financial elements “into package[s] that provides a desired combination of financing, cash-flow, and risk exposure for clients with highly particular, sophisticated financial needs” (p. 154). However, even granting Arthur that, in the ordinary course of evolution, this mechanism tends to increase complexity without reversing, it is still fairly obviously reversible through human intervention. We could, for example, simply ban the trade of certain financial assets.

<sup>179</sup> D’Agostino 2009a provides a helpful discussion.

complexity of our society is a changing rather than fixed variable, and one that we may sometimes intentionally aim to manage. If things are getting too complex, we should not just throw up our hands and give up on social reform; we should consider trying to make things less complex. This is just one more thing we have to keep in mind when attempting to solve problems of injustice or working out ways of making our society more progressive.

### *The Specific Epistemic Objection*

The general epistemic objection called into question the feasibility of even engaging in problem solving in a complex world, given my arguments against ideal target theory. I hope to have shown that this objection is unsuccessful, or at least that there are principled grounds for holding that our social world is moderately rather than highly complex, and that we can therefore consistently and plausibly maintain that even though we cannot make the sort of predictions that ideal target theory requires, we can nevertheless make the sort of predictions required by problem solving. But the critic of the progressiveness approach may remain unsatisfied, and press the following more specific epistemic objection. Even if I am right that problem solving is epistemically feasible, the second and more distinctive basic task of the progressiveness approach, theorizing about enhancing the progressiveness of our overarching framework, might not be. In particular, if it is epistemically infeasible for us to determine how to make progress toward the ideal, then why isn't it similarly infeasible for us to determine how to make progress toward an ideally progressive society? Why isn't theorizing about progressiveness just as epistemically overdemanding as ideal target theory?

This objection is important to address because it provides me with an opportunity to clarify my position. My suggestion is not that we should identify the ideally progressive society so that we can trade off short-term improvements in justice against progress toward

this progressive ideal. It is rather that we should identify short-term improvements in progressiveness, so that we can trade these off against short-term improvements in justice. In so doing, we avoid the epistemic excesses of ideal theory, because we may adopt the same orientation as problem solvers: aiming to identify and solve problems that undermine not the justice of our society, but its progressiveness. I will have much more to say about this in the next section, but the basic idea is that we may identify relatively small, short-term changes to our society that improve its capacity to experiment, learn, and select. For example, we might attempt to identify and mitigate biases that feed into our current learning or selection mechanisms. Or we might attempt to devise ways to avoid dogmatism and resistance to experimentalism more generally. Or we might identify certain phenomena, such as political polarization, that undermine progressiveness in a number of ways, and to come up with reforms that aim to remedy them.

Of course, just as ideal target theorists argue that problem solving may set back the achievement of the ideally just society, there remains the risk that, in only considering short-term improvements in progressiveness in this way, we end up recommending changes that would constitute progress away from the ideally progressive society. But, as with the worry about increasing complexity, the severity of this problem is mitigated in three ways. First, the risk itself is considerably less than in the case of pursuing ideal justice, given that short-term improvements in progressiveness generally enhance our understanding of how social features interact and our ability to make further positive changes to our society, and so generally further rather than set back our ability to continue to improve our progressiveness going forward. As already mentioned, both experimental democrats such as Knight and Johnson and proponents of experimentalist governance such as Sabel and Zeitlin emphasize that progressive institutions must involve some element of reflexivity—some ability to

experiment, learn, and select not only with reforms aimed at improving justice, but with reforms aimed at improving progressiveness—if they are to improve or at least sustain their progressiveness in our ever-changing world. So, in general, we should expect that promising ways of improving progressiveness will correlate with promising ways of improving our ability to improve progressiveness, even though it would be too optimistic simply to assume that this will always be the case.

Second, as we saw when discussing the moving target objection (in Chapter 3), it is implausible to think that there is a fixed ideally just or ideally progressive social arrangement that we could cut ourselves off from in the first place, given that changes to technology, the environment, demographics, and so on, interact with institutions to affect their justice or progressiveness, and it is implausible that such changes will ever stop. The risk is therefore not so much that we might cut ourselves off from a fixed progressive ideal, but that we might implement reforms that more generally set back the achievement of greater justice or progressiveness in the long term. But put in these terms, the argument of this chapter has been precisely that focusing on trying to make relatively small changes to our society that improve our progressiveness is the best way to reduce this risk given our epistemic situation, even if it cannot remove this risk altogether.

Thus, third, and closely relatedly, I do not rest my argument on the claim that changing our social arrangement in ways we predict will improve progressiveness is in all cases guaranteed to maximally promote long-term justice. In a complex world, no such guarantee is available. Instead, my argument rests on the comparative claim that trading off short-term improvements in justice against short-term improvements in progressiveness is a better approach to pursuing social reform than either focusing only on improving short-term justice through problem solving or attempting to trade off short-term justice against making

progress toward the ideal. The progressiveness approach therefore represents the appropriate middle ground between the myopia of the former approach and the impracticability of the latter. It threads the needle.

#### 4.5. Applying the Progressiveness Approach

The progressiveness approach, then, appears to have fairly strong epistemic credentials, and so appears to be our best hope for theorizing about social reform in a way that accommodates both a concern with long-term justice and our epistemic limitations. But one important desideratum of any new approach is that it is fruitful, and since we have not yet attempted to apply the approach, we have not yet seen whether the approach will bear much fruit. Here, I cannot fully satisfy those who fear that it will not; my hope is that many other theorists will take up the task of working out how to improve progressiveness, and the upshot of such work remains to be determined. But I nevertheless believe there are good reasons to have high hopes for the progressiveness approach, and to show this, I will now briefly discuss how one might go about applying it. Since my goal here is to inspire further work on this topic rather than to carefully defend particular ways we should improve progressiveness, I will, in what follows, sacrifice depth for breadth. Nearly everything I will go on to say in this section is subject to objection, but that is okay. Here, I am just trying to get the conversation going, and to show that there really is a conversation to be had.

Even with this caveat stated, we run into an immediate complication when attempting to apply the progressiveness approach. Throughout this essay I have simply set aside the issue of what criterion of justice to use when evaluating social arrangements by the outcomes they produce, and have asked the reader to apply her favored criterion to the issues at hand. But theorizing about progressiveness might seem to require us now to

endorse some such criterion. It might seem to require, in other words, that we first adopt some particular criterion of what makes for a more or less just society, so that we can then work out ways of institutionalizing mechanisms that make our society faster and more reliable at climbing higher and higher on that criterion across time.

It is certainly possible to theorize about progressiveness in this sort of *criterion-dependent* way. Suppose, for illustration's sake, that one endorses a simple, Rawlsian-inspired criterion of justice, according to which one society is more just than another when it is better for the worst off. Given this criterion, our efforts at improving progressiveness should focus on making our society faster and more reliable at improving the prospects of the worst off. So, taking inspiration from experimental democrats, we might attempt to make our voting mechanisms more responsive to the interests of the worst off, such that, over time, the democratic process does a better job of implementing and stabilizing social arrangements that are better for the worst off and at destabilizing those that aren't. Or, drawing on the literature on polycentricity, we might focus on trying to work out policies that provide the worst off members in our society easier access to information about their prospects in other jurisdictions, as well as the ability to move from jurisdiction to jurisdiction at low cost, so that those who find themselves in jurisdictions that are worse for their prospects may travel to those which are better for them. Or we might follow proponents of experimentalist governance in attempting to establish the improvement of the conditions of the worst off as a broad framework goal. To recall: this would involve implementing provisional standards for measuring the interests of the worst off, providing local administrative units with autonomy to experiment with different ways of trying to better the worst off on the condition that they report the results of their efforts in processes of peer review, and establishing methods of reviewing our provisional standards to ensure that agencies are not

just improving across some proxy measure but genuinely bettering the worst off.

Of course, complications arise for any of these approaches. For instance, given the need to experiment in order to expand our understanding of how to benefit the worst off, experimental democrats would not want to make our democratic institutions so good at selecting for or stabilizing what, given our current information, seems in the best interests of the worst off, that they no longer leave room for experimentation with novel reforms. But this is more of a worry in theory than in practice, since it is patently unrealistic to think we will ever get to the point where our democratic system is really just selecting policies that our best evidence suggests favors the worst off: in a diverse society characterized by widespread disagreement in interests, criteria of justice, and empirical beliefs about what would best satisfy people's interests or criteria of justice, exploring other arrangements via the democratic process comes, in a sense, for free. Similarly, we might worry about both the feasibility and desirability of the suggestion that we simply make it much easier for the worst off to move to jurisdictions in which they are better off. How would we do this? And even if we could, what would prevent jurisdictions from simply ignoring the interests of the worst off, saying "good riddance" to them as they head off to other jurisdictions, leaving the worst off with no good options at all? Presumably, we would either need to provide jurisdictions with strong incentives for housing the worst off members of society, or to employ this polycentric solution concurrently with either the experimental democratic or experimentalist governance approach to prevent such issues from arising. Finally, with respect to the last approach, can we really expect all administrative agencies to simply adopt the promotion of the worst off as a framework goal? And would this even be desirable given the benefits of specialization that come from having administrative agencies pursue more specific goals? The trick would likely be to have administrative agencies pursue a range of more particular

policy goals whose effective pursuit would strongly correlate with improving the position of the worst off, but it is unclear how we would pull this off.

All of this requires much further discussion before we should feel confident recommending any concrete progressiveness-enhancing reforms. And, furthermore, which particular reforms we would want to implement depends on our criterion of justice: different mechanisms may be more appropriate to trying to make our arrangements faster and more reliable, say, at improving total well-being, better protecting certain rights, more equally distributing resources, reducing racial and gender discrimination and oppression, or so on, than at increasing the prospects of the worst off. But rather than focusing further on how we might try to implement such reforms that are deemed progressiveness-enhancing given some particular criterion of justice, I now want to shift focus and ask: are there certain reforms that are *criterion-independent*, in the sense that they should be able to garner broad support from individuals with a wide range of conceptions of justice? Such criterion-independent reforms are especially important for theorists of progressiveness to focus on, since by their very nature they are much more likely actually to be implemented than those whose appeal one can only recognize given one's endorsement of some particular criterion of justice. They are also more likely actually to improve progressiveness, since unlike criterion-dependent reforms, their desirability is not so contingent upon the presumed accuracy of the particular criterion of justice on which they depend.

On first glance, there might seem to be an in principle difficulty with coming up with criterion-independent reforms. If you and I disagree about what criterion explains when one society is more just than another, then it might seem hard to imagine how we can agree about what makes one society faster and more reliable at improving in justice. But there are a number of reasons to think that this difficulty is overstated. First, even if we disagree about

fundamental criteria of justice, we might agree about more specific goals. A wide range of theories of justice can agree that it is important to protect certain rights, to ameliorate poverty, to decrease income inequality, to reduce racial and gender discrimination, and so on, even if they do not agree about the explanation of why these things are important or what their weights are when these goals come into conflict with one another. Second, even given residual disagreements over specific goals and their weights, we might agree about the relative justice of particular outcomes. Perhaps you and I disagree about why the successes of the civil rights movement created a more just society in the United States, but we both agree that it did. Or perhaps we both agree that the consequences of rising economic inequality undermine justice, but not about which particular consequences are relevant.<sup>180</sup>

Third, and most importantly in the current context, even if we disagree about the ranking of a significant range of particular outcomes, we might still reach agreement about the sorts of factors that tend to produce more or less just outcome over the long term. For example, even if we do not agree about what criterion of justice our selection mechanisms should be aiming at, we might both agree that, in general, we tend to achieve less just

<sup>180</sup> Our ability to agree about particular goals, outcomes, norms, or institutions even despite different underlying conceptions of justice or morality is emphasized by proponents of “convergence” approaches to public reason liberalism, such as Gaus 2011 and Vallier 2014. Although it may be difficult to believe that we would converge in this way on what the ideal society would be like, such theorist stress that it is often easier to find agreement about the ranking of particular options. Gaus and Vallier focus primarily on cases where we can agree that the existence of some norm or institution is preferable to its absence, but more germane for our purposes are cases where we are able to reach agreement about whether some change would represent an improvement over the status quo.

outcomes when the political process is heavily biased toward the interests and opinions of the rich and powerful than when it is more responsive to the public at large. We might, in other words, agree about certain ways we might debias our selection mechanisms or more generally make them better correlate with each of our understandings of justice, even if we do not agree about what a perfect correlation would look like. Or we might agree that the more information we are able to generate and widely spread throughout society about how social features interact to produce their effects, the more likely we are to achieve more just outcomes, since we each hold that our circumstances are such that, even though this information will not always be used in ways that favor justice as each of us understands it, this information will at least be used more often to promote justice than to set it back.

In principle, then, we may be able to reach widespread agreement on reforms that are progressiveness-enhancing, even if we disagree about our criteria of justice. Of course, the particular reforms we will want to implement will still differ depending on the particular society we are attempting to reform. But, in general, there are certain things we might want to watch out for. For example, we may sometimes find that while our society is trending in the right direction given a wide range of plausible criteria of justice, it is not improving very quickly, because it is stagnant, insufficiently able to engage in a wide range of promising experiments. In these cases, progressiveness may be improved in a number of ways. For example, we might allow for more diachronic experimentation by attaching more “sunset clauses” (set end-dates) to policies, by making greater use of pilot projects, or by allowing for more concurrent experimentation through providing greater autonomy to jurisdictions or administrative agencies to try out their own reforms.<sup>181</sup> Or we might similarly incentivize or

---

<sup>181</sup> Müller 2019, pp. 147-154, canvasses a number of different ways we might allow for more concurrent experimentation via the establishment of what he calls a polycentric democracy.

at least facilitate (say, via better information pooling) greater cross-jurisdictional and -national copying and recombination of social features, given the crucial role of recombination of simpler elements in driving innovation nearly across the board.<sup>182</sup>

To be sure, there are significant short-term justice costs associated with engaging in certain experiments. Since the goal is to implement mechanisms that permit a wide range of promising but not impermissible or otherwise unpromising experiments, a progressive society will likely be one that takes some options off the table, say, because they do not accord with a bill of rights. But even when it comes to promising experiments, too rapidly experimenting with different social arrangements may itself undermine short-term justice. Excessive experimentation may be intrinsically unjust because it undermines legitimate expectations, or unjust because its unsettling of expectations has further negative effects, say, on individuals' ability to anticipate the features of the social world they will encounter and so to plan accordingly.<sup>183</sup> Indeed, even apart from such short-term justice costs, overly rapid experimentation may even undermine our capacity to learn, since, in order to know what

---

<sup>182</sup> The role of recombination is most obvious with technology—for example, “[t]he laser printer was constructed from the existing laser, digital processor, and xerography” and “the railway locomotive was constructed from the already existing steam engine, boiler, cranks, and iron wheels” (Arthur 2015, pp. 19-20)—but recombination of simpler elements is essential to producing innovation in all complex systems, since it helps us to discover and put to work promising “building blocks” or “schemata.” See especially Holland 1992. Indeed, on Holland’s analysis, it is recombination (via sexual reproduction), rather than random mutation, that drives much genetic evolution, though mutation remains essential since it “assures that no allele permanently disappears from the population” (Ibid., p. 110).

<sup>183</sup> See Simmons 2010, pp. 20-21, and, for a classic discussion Sidgwick 1981, pp. 268-273.

effects certain combinations of features have, we often need to wait a while, and let the interactions play out. So in thinking about how to make our society more progressive, we should not assume that we want to maximize the number of experiments we are engaged in; as mentioned earlier, we need to trade this off not only against short-term justice but also against their informativeness. Still, there are various ways we might try to trike this balance. For example, instead of using simpler sunset clauses that mandate that we reconsider policies after a certain number of years, we might implement clauses that also set a high procedural bar to revising those policies before those years are up, which are passed along the following lines: “This is a serious problem. We propose to initiate Policy A on an experimental basis. If after five years there has been no significant improvement, we will shift to Policy B.”<sup>184</sup>

In the end, however, perhaps nothing will impact our ability to engage in promising, informative experiments more than changing our informal norms. As John Dewey famously quips: “Men have got used to an experimental method in physical and technical matters. They are still afraid of it in human concerns.”<sup>185</sup> Although the advent of the evidence-based policy movement and the growing use of experimental methods in attempts to solve problems such as global poverty suggest that this may have changed somewhat since Dewey’s day, if the state of contemporary political debate is any indication, there is still much room for improvement.<sup>186</sup> Just think, for instance, of our norms surrounding the evaluation of politicians. We tend to see it as a bad a thing that a politician is a “flip-flopper”

---

<sup>184</sup> Campbell 1969, p. 2.

<sup>185</sup> Dewey 1927, p. 169.

<sup>186</sup> See Banerjee and Duflo 2012 for an overview of their attempts to mitigate global poverty through the use of randomized controlled trials. This work earned them the 2019 Nobel Prize in Economics “for their experimental approach to alleviating global poverty.”

who changes her policy platform over time, as if a good politician is someone who should remain ideologically committed to particular policies through thick and thin. But perhaps this is a mistake: if we want a progressive society in which we are able to experiment with a wide range of solutions to the problems we face, then we may instead want our politicians to be people who are able to come up with clever and innovative ways to experiment with and monitor solutions to problems, but who are also willing to change their mind in response to new evidence as it arrives. Indeed, such alternative norms for evaluating politicians would have the added benefit that they would make us more likely to elect the very sort of people that we know from Tetlock's work are better at making political predictions. Again and again, we find that the best forecasters are people who are willing to second-guess themselves, consider issues from a wide range of perspectives, and revise their views in the face of new evidence, and that the worst forecasters are those who are highly ideological (no matter their particular ideology), begin very confident in their world views and predictions, and prove unwilling to revise them in response to new evidence.<sup>187</sup>

---

<sup>187</sup> This is the basic lesson of Tetlock 2006, which is neatly summarized at Tetlock and Gardner 2016, pp. 55-56. Remarkably, Tetlock 2006, p. 68, finds that predictive capability correlates negatively with fame. As Tetlock and Gardner 2016, pp. 57-58, explain, this is because bad forecasters tend to employ a "one-perspective analysis" that allows them to "pile up reasons why they are right—"furthermore," "moreover"—without considering other perspectives and the pesky doubts and caveats they raise"; this allows them to deliver simple messages that resonate better than those delivered by more adept forecasters who look at problems from many angles and so employ various "“howevers” and “on the other hands”" (pp. 57-58). This is a correlation that a change to our informal norms could hopefully reverse, though it is hard to know how much it depends on fixed psychological tendencies.

Revising our norms relating to our evaluation of politicians and, for that matter, administrators, policy advisors, and members of the media would also plausibly help us better to learn and select. Among other things, it would help us to overcome a common problem in contemporary societies that Campbell calls the “overadvocacy trap.”<sup>188</sup> In brief: the trouble is that, in many contemporary societies, political experts have an incentive to oversell the benefits of their proposed reforms in order to increase their chance of being passed, and then a further incentive to prevent the monitoring of these reforms because they know that they are unlikely to live up to their bill of goods. But this clearly interferes with both learning and selection: if we aren’t monitoring the effects of our reforms, we can neither learn from them nor destabilize them if they turn out not to work. Changing the way we evaluate political actors may help with this, because if they know that their job security and reputation depends on their demonstrated ability to cleverly experiment with various policy proposals and to revise their suggestions in light of new evidence, rather than on their ability to make it seem like they successfully guessed what would work ahead of time, their incentive to avoid monitoring their proposed reforms would disappear.

Another potential solution to the overadvocacy trap, that would also more generally help us to learn, would be to make greater use of formal monitoring, perhaps even requiring that all political policies come with built in measurement standards and monitoring mechanisms that permit us to see how well they are living up to expectations, and to learn from deviations from expectation.<sup>189</sup> Of course, a well-known problem with the

---

<sup>188</sup> Campbell 1969, 1973.

<sup>189</sup> In complex systems, we are generally better at learning by examining such deviations in expectation than from trying to examine all the factors producing an outcome, since we can rarely canvas all such factors. As Axelrod and Cohen 2000, p. 143, explain: “The search is

establishment of such measures is that they may lead to a situation in which we are improving across such measures, without actually improving along the dimensions the measures are intended to track.<sup>190</sup> But there are various ways to mitigate this concern. For example, we might employ a variety of measures for any given justice-relevant variable we are concerned with rather than relying on a single measure. And we might recursively update measurement standards in consultation with a wide range of relevant stakeholders by employing techniques like the “snowball approach,” whereby those tasked with setting up and revising measurement standards not only ask all the stakeholders that they can think of for input, but also ask those stakeholders who else they think they should be consulting with, and so on, until, hopefully, all relevant individuals are consulted.<sup>191</sup>

In addition to attempting to overcome the overadvocacy trap in these ways, there are a number of things we might try to do improve a society’s learning mechanisms more generally. For example, the snowball approach and related techniques might be used more widely to attempt to debias our learning mechanisms through mitigating the effects of epistemic injustice—here understood as the unfortunate tendency to give greater weight to the opinions of members of dominant social groups than members of marginalized groups when monitoring existing arrangements and looking for problems.<sup>192</sup> To eliminate the root cause of epistemic injustice, however, we may need to reform things further upstream, for 

---

not for what predicted the outcome but for what predicted the surprise, the deviation of your expectations from what occurred. Those are the factors to which you should give increasing credit if you want to speed the process of learning which factors to credit.”

<sup>190</sup> For classic statements of this worry, see Campbell 1979 and Goodhart 1984.

<sup>191</sup> Rossi, Lipsey, and Freeman 2003, pp. 87, 124.

<sup>192</sup> Fricker 2007 provides a seminal discussion of epistemic injustice.

example, by reducing segregation between such groups in educational institutions and communities of inquiry more generally.<sup>193</sup>

Learning is also impeded by the cluster of phenomena—including the fractured media environment, political polarization, and the advent of fake news—that characterize the current environment in countries such as the United States. If different groups trust different news sources, some of these news sources are not very trustworthy, yet individuals do not trust the testimony of those in other groups and so cannot correct one another's errors, then individuals cannot learn from each other and information cannot flow in productive ways. Obviously, these are big problems with no easy solutions, but various proposals have been suggested about how we might try to regulate fake news and more generally improve our media environment, perhaps through identifying and appropriately regulating market failures in the “marketplace of ideas,”<sup>194</sup> as well as how we might reduce polarization, for example, through implementing reforms that promote cross-cutting identities or cleavages between currently segregated groups.<sup>195</sup> A related but more general worry, bearing on both our ability to learn and to select through democratic means, is that voters may be woefully ignorant about political matters because, given the vanishingly small chance any individual's vote makes a difference, voters have hardly any incentive to gather

---

<sup>193</sup> Anderson 2012.

<sup>194</sup> See, for example, O'Connor and Weatherson 2019, ch. 4.

<sup>195</sup> For a recent treatment, see Klein 2020, who argues that polarization has risen sharply in the U.S. today because individuals have become far more sorted into two distinct camps along party, ideological, and demographic lines than they used to be. In the past, one could find a variety of ideologies and demographic groups cutting across party lines; now, each party is relatively homogenous.

such information.<sup>196</sup> This raises a number of complicated issues in democratic theory, so here I will just point out that if the problem is that individuals lack sufficient incentives to get informed, then we should begin to think about different ways of generating those incentives—from financial incentives, to providing individuals with a greater ability either to move between jurisdictions (so that they have a stronger incentive to learn what is happening in each)<sup>197</sup> or to participate in workplace governance (which may require them to learn similar information).<sup>198</sup> Education may also play an important role here as well.

Finally, focusing more squarely on selection, we might experiment with various ways of debiasing existing democratic selection mechanisms to make them less responsive to a select group of the rich and powerful through such means as campaign finance reform or the greater use of more directly democratic institutions such as deliberative councils and citizens' assemblies.<sup>199</sup> And then, as already mentioned, there are various ways we might try to improve selection via competition between jurisdictions by, for example, reducing exit costs between such jurisdictions, or to improve selection via administrative discretion by revising our administrative agencies along experimentalist governance lines. Further, while my focus

---

<sup>196</sup> See, for example, Brennan 2017 for an influential argument against democracy drawing on this sort of concern, and Christiano 2019 for a reply to this style of argument, challenging both the premise (that voters lack the necessary information to make good political decisions) and the conclusion (that democracies do not produce good outcomes as a result).

<sup>197</sup> Somin 2010.

<sup>198</sup> Christiano 2019.

<sup>199</sup> For some evidence concerning the existence of this bias, see Bartels 2017. Forstenzer 2019, ch. 10 provides a helpful overview of various proposed “democratic innovations” that may help to solve this problem from the perspective of democratic experimentalism.

has been on individual societies, there is also an important role here for selection through global governance involving international organizations such as the UN Security Council and the International Criminal Court who have some ability to penalize societies for departures from widely accepted standards of justice as enshrined, for example, in international conventions on human rights.<sup>200</sup> Again, figuring out criterion-independent ways of improving selection mechanisms is always difficult given widespread diversity in the criteria of justice and morality that people endorse, but this is possible so long as we are nevertheless able to agree about more concrete political policy goals, or at least about certain ways in which our current selection mechanisms tend to go wrong. Indeed, there are reasons to think that diversity, so far from being a hindrance to progressiveness, may actually be a major driver of progress, given a wide range of theoretical and empirical evidence suggesting that diverse groups of individuals tend to be better at both predicting and at solving problems even than groups of people who are individually more capable.<sup>201</sup>

---

<sup>200</sup> See Buchanan and Keohane 2006, who, though focused on legitimacy rather than justice, provide a helpful discussion of the epistemic virtues we should demand of such international organizations, involving their ability to channel information in productive ways, to meet certain standards of accountability, and to revise these standards across time.

<sup>201</sup> See especially Page 2007 and Landemore 2012. However, we need to be careful not to oversell the benefits of diversity here, for two reasons. First, the capacity of diverse groups to make better predictions and to better solve problems depends on their inclusion, respectively, of a diversity of predictive models and cognitive styles, both of which only imperfectly correlate with other forms of diversity. Second, while diversity in individuals' goals or evaluative standards is beneficial insofar as it does in fact correlate with predictive and cognitive diversity, too much evaluative diversity can undercut these benefits by, among

Now, all of this has been rather surface-level, and is in no way meant to be a thorough canvassing of ways we might improve progressiveness. But let me reiterate that my purpose here has been only to gesture toward some promising future directions for research about how to promote progressiveness, and to show that there is indeed much more research to be done. If I have convinced the reader of that much—even if they are less satisfied by my more specific suggestions—then I consider this section a success.

Still, before closing this section, let me note one thing that, I think, we can be fairly confident would improve progressiveness—namely, that we must shift the practice of political philosophy from its obsession with ideal theory to an embrace of the progressiveness approach. This shift, I submit, would itself improve progressiveness for two major reasons. First, the more political philosophers, theorists, and so on, engage in problem solving and theorizing about progressiveness, the more likely they are to come up with reforms that would genuinely improve justice and progressiveness, and to inspire others to do the same. Obviously, there is a big gap between crafting a persuasive defense of a reform and actually having that reform implemented or emulated, but, throughout history, nearly all major reform movements been spurred by the development of an intellectual movement.<sup>202</sup>

---

other things, driving up communication costs. Page 2007 is crystal clear about all this, and very thorough. See also Gaus 2018, ch. 3, for helpful discussion, including some suggestions about how we might structure society in order to harness the benefits of diversity without running into such costs, as well as Muldoon 2016 and Müller 2019 for approaches that emphasize the importance of both experimentation and diversity to social progress.

<sup>202</sup> See Rochon 1998 for an argument to this effect drawing on the history of a wide range of social movements, including the French revolution, the movement to abolish slavery, the civil rights movement, the women’s liberation movement, and the environmental movement.

So the more wind the progressiveness approach has in its sails, the more likely the work of those thinkers aiming to improve progressiveness is to have uptake. Thinking this does not commit me to John Maynard Keynes's famous claim that "the world is ruled by little else" than "the ideas of economist and political philosophers."<sup>203</sup> It commits me only to the much weaker claim that such ideas have enough power to at least make some genuine difference to the course of social change.

Second, I have suggested that one major way we might improve progressiveness is to shift our informal norms, and especially our norms for evaluating political experts, in the direction of experimentation rather than dogmatism. We need to change the way people think about and discuss politics, from an attitude where we take serious political thinkers and leaders to be those who stand unwavering in their ideological commitments to concrete long-term goals, to one in which we recognize that it is very difficult to know where we should be aiming at in the long-term, such that we must instead be focusing on experimenting with a wide range of solutions to the problems we face, on working out ways of learning from and improving our ability to experiment with and learn from those reforms, and on trying to improve our selection mechanisms so that better experiments tend to stabilize whereas worse ones don't. We must, in other words, shift people from a sort of implicit endorsement of the ideal theoretic approach, to an acceptance of something more like the progressiveness approach. So the change I am advocating for must occur not only in academic political philosophy or theory, but also in the way that politicians, administrators, educators, activists, members of the media, and the general public think about, and expect one another to think about, politics in general and social reform in particular.

Of course, simply adopting the progressiveness approach in academic political

---

<sup>203</sup> Keynes 2018, p. 340.

philosophy would not be sufficient to change these wider norms. But it would certainly help. Political philosophers influence the general public in various ways, from being called on to comment on political events, to writing amicus briefs that influence the decisions made by the Supreme Court, to publishing works in academic journals and presses as well as in more popular venues, to educating future leaders and present citizens. We, as a society, need to change the way we approach questions of social reform, so that we can better achieve greater justice over the long run. And we, as political philosophers, should be at the forefront of this movement, not lagging behind it. That requires as many of us as possible to embrace the progressiveness approach.

## CHAPTER 5 - CONCLUSION

## 5.1. The Mountain Analogy

I began this essay by outlining two distinct but related debates going on in political philosophy. One concerns the practical relevance of ideal theory, and, in particular, the relevance of theorizing about the ideally just society to questions of social reform. The other is a debate over the extent to which political philosophy must be practically relevant in the first place. To this point, I have focused on the first debate, and on the related question of how we should engage in reform theory without recourse to ideal theory. So let me first sum up my discussion of the former debate before turning, at last, to the latter.

Within the debate over the practical relevance of ideal theory, we have seen that there are two broad views about why a conception of the ideally just society is practically relevant. There is the ideal benchmark view, according to which the ideal serves as an indispensable benchmark for evaluating societies as more or less just. And there is the ideal target view, which claims that the ideal serves as a long-term target for reform. In Chapters 2 and 3, I argued that each of these views is mistaken. Although a conception of the ideally just society might sometimes help us to theorize about how to improve our society—it may help us to develop our criteria of comparative justice, to generate novel solutions, and to appreciate present problems of injustice—it plays no indispensable or even central role in this regard. And the ideal cannot serve as a long-term target for reform both because we lack the epistemic capacity to identify it and work out how to pursue it with sufficient confidence to warrant this pursuit, and because there is likely no fixed ideal to pursue anyway.

Despite these failures of ideal theory, I have nevertheless acknowledged that ideal target theorists are right to point out a defect in the other major approach to reform theory on offer: the problem solving approach. When we engage in problem solving, we attempt to

identify concrete instances of injustice in our society, and to come up with narrowly targeted solutions. But, as we saw in Chapter 3, while problem solving is a good way to theorize about short-term justice, it is not a good way to theorize about long-term justice, because the targeted solutions we implement may interact either to undermine overall justice or to set back the prospects of future improvements. So, in Chapter 4, I proposed a new way to theorize about long-term justice. In particular, I argued that we should supplement problem solving not with ideal theory, but with an attempt to work out how to make our society more progressive: better at getting better, or more conducive to further improvements in general. And I argued that, to do this, we must focus more specifically on trying to improve our society's ability to flexibly experiment with many promising solutions to problems of injustice as they arise, to select for those that prove successful while eliminating those that do not, and to help us learn from both our successes and our inevitable failures.

The various arguments I have presented against problem solving and ideal theory, and in favor of the progressiveness approach, have been rather complicated, drawing sometimes on fairly abstract claims about axiology, though more often on considerations of social complexity. But the basic sketch of the argument is, I think, quite simple, as can be illustrated by way of an analogy. In the literature on the ideal/nonideal theory debate, it is common to compare social reform to mountain climbing. The analogy derives from Sen's critique of the benchmark view, in which—as we saw in Chapter 2—Sen compares making comparative evaluations of justice to making comparative evaluations of the heights of mountains.<sup>204</sup> From here, the idea is that pursuing justice is like traversing a mountain range: how high we have climbed corresponds to how just our society has become. Now, if we care only about getting higher (or improving justice) in the short-term, then we can set off

---

<sup>204</sup> Sen 2009, p. 102.

without having any particular end-goal in mind: we can simply always move upward (or toward greater justice) from where we are, without worrying about whether we are ultimately moving away from the highest point (or ideal justice). This is basically what problem solving recommends, at least in its simplest form: we try to figure out particular ways of making our society more just without worrying about the further consequences. But ideal target theorists argue that if we do eventually want to reach the summit (or the ideal), then, we better not set out “blindly,” without any idea where to find it. For, when it comes to deciding which transitions to pursue, “which of two smaller ‘peaks’ of justice is the higher (or more just) is a judgment that matters conclusively only if they are both on equally feasible paths to the highest peak of perfect justice. And in order to endorse a route to that highest peak, we certainly *do* need to know which one that highest peak is.”<sup>205</sup>

By now, we have seen that this analogy is rather misleading. The problem is not, as ideal benchmark theorists might claim, that whereas we don’t need to identify the highest mountain to apply a criterion of comparative height, we do need to identify a conception of the ideally just society before we can apply criteria of comparative justice. As I have argued, plausible criteria of comparative justice really are like criteria of comparative height, at least in the following sense: they are essentially comparative and so can be applied without making any reference to the ideal. Instead, the problem with the mountain analogy is that it ignores the epistemic differences between mountain climbing and pursuing social reform. After all, we already know that Mount Everest is the highest mountain, and mountain climbers are fairly adept at tracking their progress toward it. They generally know, from

---

<sup>205</sup> Simmons 2010, p. 35. Although Simmons puts this point in terms of making the ideal more feasible, it can be easily translated into our more familiar talk of making “progress toward” it.

where they are on the mountain range, which direction they need to head off in if they wish to increase the odds they will ever reach the summit or decrease the time that will take. But, as I have argued, we lack the corresponding knowledge in the case of justice: we can neither identify the ideal, nor track our progress toward it. We are therefore more like mountain climbers who have no idea where the highest point is and have no real prospect of figuring out which paths lead us toward it, but who do still want to get as high as possible. More specifically, since in the case of justice we must trade off the pursuit of short-term justice against the pursuit of long-term justice, we are like climbers who care both about getting higher in the short term, and about getting higher in the long term, but who are exploring a largely uncharted mountain range with little understanding of the lay of the land. The tradeoff we face is therefore not between getting higher in the short-term and making progress toward the summit, but, rather between getting higher in the short-term and improving our prospects of getting higher in the long-term. And this is the same way I have suggested we must conceptualize the trade off in the case of justice: we must trade off short-term justice not against making progress toward ideal justice, but against becoming more progressive, or getting better at improving in justice in general.

Now, in the case of mountain climbing, there are a number of things we must do to try to get higher in the long-term. For one, we must be able to explore, and more generally to gather information about, the landscape. So we should, for example, focus not only on working out how to head upward from where we are, but also on taking steps that will teach us more about the surrounding area. Better yet, we should work together with a team of mountain climbers and scouts, each exploring different regions of the landscape at the same time, and transmitting this new information back to other members of the group (say, over walkie talkies). But how much any given climber will learn from a given point on the

mountain range, and her more general ability to move about in ways that help her to get higher in both the short and the long term, depends on certain more general facts about the climber's abilities. Her capacity to learn and retain information, for example, might be improved by her acquisition of binoculars or a pen and paper on which to draw a map. And her ability to avoid getting stuck, or more generally to move nimbly from point to point, may depend on her physical fitness, her technique, or on what sort of gear she has.

There are, of course, clear disanalogies between mountain climbing and social reform; it would be tedious to list them all here. But I nevertheless think that the analogy is illuminating because it illustrates that it is possible to trade off short-term improvements against long-term improvements even without having any particular long-term goal in mind. And it also helps us to see the rough idea behind my claim that experimentation, learning, and selection are key to improving progressiveness. Just as our mountain climbers must be willing to explore physical space, preferably by engaging in a wide range of simultaneous excursions, progressive societies must be able to explore the space of social possibilities by engaging in a wide range of social experimentation, preferably not only diachronically but simultaneously. Just like our mountain climbers are better positioned to reach greater heights in the long run if they have binoculars, the ability to draw a map, and the ability to transmit information to one another over the radio waves, our society is more progressive if it has better learning mechanisms in place. And just as our mountain climber's ability to climb upward and to avoid slipping back downward depends on her physical prowess, technique, and gear, our society's progressiveness crucially depends on its selection mechanisms.

Of course, it is one thing to draw this very rough analogy, and another to say, how, precisely, we should make our society more progressive in general, and better at experimenting, learning, and selecting in particular. I began to explore how we might go

about answering these questions at the end of the last chapter, but there is much more work to be done. So those of us who aspire to doing practically relevant political philosophy should begin this work posthaste.

But what if one doesn't aspire to this? What if one has a purely intellectual interest in justice, and doesn't care much about whether one's work is practically relevant or has good practical effects? Should one prefer the progressiveness approach to ideal theory anyway? Yes, one should, and I will now briefly explain why.

## 5.2. Anti-Practicalism Revisited

If practicalism is the view that the value of an instance or type of political philosophy depends entirely on its practical relevance or its practical effects, then anti-practicalism is the denial of this view. According to anti-practicalists, it matters not only that our theorizing is relevant to the promotion of justice, but also that we acquire true beliefs or knowledge about justice, or that we come to better understand justice.<sup>206</sup> Although it might seem like my arguments so far should only convince practicalists to abandon the ideal theoretic approach for the progressiveness approach, I now wish to show that anti-practicalists should also prefer the progressiveness approach. This is because the progressiveness approach dominates the ideal theoretic approach, not merely as an approach to reform theory, but also as an approach to theorizing about justice more generally: it does better not only with respect to practical value, but with respect to intellectual value, too.

That the progressiveness approach has greater practical value than the ideal theoretic approach follows straightforwardly from what I have already argued. Ideal theory is only peripherally relevant to theorizing about reform, whereas both problem solving and

---

<sup>206</sup> Estlund 2011a, 2011b, 2014; Cohen 2008.

theorizing about progressiveness are directly concerned with it, and these are the two basic tasks recommended by the progressiveness approach. Furthermore, I argued at the end of Chapter 4 that were political philosophers to embrace the progressiveness approach, this itself would contribute to a more progressive society. So the widespread adoption of the progressiveness approach would not only lead us to address more practically relevant questions, but it would also likely produce better practical effects as well.

Perhaps more surprisingly, however, the progressiveness approach also provides greater intellectual value than the ideal theoretic approach. Ideal theorists, recall, aim either to describe fairly abstract principles that would be realized by the ideally just social arrangement, or to describe the concrete features of that arrangement. But, beginning with the latter, we have seen that we simply lack the epistemic wherewithal to describe the ideal society in any concrete detail with very much confidence at all. Ideal theorist might arrive at lots of false beliefs about this and might stumble across the occasional true one. But they are unable to get well-justified beliefs, knowledge, or understanding about what the ideal social arrangement would look like. I have argued, however, that we do have the correlative epistemic capacity to work out how to make our current social arrangements more just or more progressive. So engaging in problem solving or theorizing about progressiveness would permits more true beliefs, knowledge, and understanding about justice, than theorizing about the ideal social arrangement. Furthermore, if adopting the progressiveness approach would indeed help to make our society more progressive, this would tend to bring with it even more truth, knowledge, and understanding over the long run, since one feature of a progressive society is that, within it, we are better capable of learning about justice.

This leaves only the possibility that it is ideal theorists' focus on the abstract principles that would be realized in the ideal society that provides ideal theory with more

intellectual value than the progressiveness approach. And, indeed, I have not presented any epistemic argument against our ability to accurately characterize the ideally just society when described in such abstract terms. Even still, there is no reason to think that the sort of abstract theorizing that is characteristic of ideal theory has any greater intellectual value than the correlative form of abstract theorizing that, as I mentioned in the last chapter, makes up an important (though subsidiary) task of the progressiveness approach: theorizing about abstract criteria of what makes a society more or less just. Indeed, I have argued at length in Chapter 2 that criteria of comparative justice are essentially comparative, such that the ideal should simply be understood as the society that is more just than all other societies (given the correct essentially comparative criterion or criteria) in some relevant constraint set. For example, if the correct criterion of comparative justice says that one society is more just than another when it is better for the worst off, then the ideal society is simply the society that is better for the worst off than all other societies meeting certain constraints. So the only extra information we need to describe the ideally just society once we have already settled on our criterion or criteria of comparative justice is which constraints are relevant—for example, whether the ideal is the society that is more just than all other logically possible societies, or the society that is more just than all other societies that are compatible, say, with certain fixed physical and psychological facts. Thus, the only truth, knowledge, or understanding, we might gain from theorizing about ideal principles that we would not similarly gain from theorizing about abstract criteria of comparative justice is which constraints are relevant to characterizing some society as ideal. And it is simply implausible to think that the intellectual value of *this* really outweighs the greater intellectual value of believing, knowing, or understanding the correct criteria of comparative justice, or believing, knowing, or understanding the various truths that we arrive at about how to improve justice or

progressiveness on which the progressiveness approach is most centrally focused.

I therefore submit that there are both strong practical and strong intellectual reasons to adopt the progressiveness approach over the ideal theoretic approach. Especially forceful among these reasons is the fact that adopting the progressiveness approach would likely make our society more progressive, and so, in the long run, a more just place. In other words, the widespread adoption of the progressiveness approach would itself constitute a promising social reform, one justified on the grounds that it makes our society more progressive, more conducive to further improvements in general. And adopting it would come with no significant costs from the perspective of either practical or intellectual value, at least among those already interested in theorizing about justice.

This completes my argument that we ought to adopt the progressiveness approach not only so that we can better understand justice, but also so that we can better pursue it. It also brings to a close my discussion of the two debates currently raging in political philosophy. In response to the debate over the practical relevance of ideal theory, I have argued over the last three chapters that problem solving and theorizing about progressiveness are far more practically relevant. And I have now more briefly addressed the second debate over the extent to which political philosophy must be practically relevant by arguing that, regardless of one's take on this latter issue, those interested in justice should be able to accrue greater truth, knowledge, or understanding about justice by adopting the progressiveness approach rather than ideal theory anyway. So anti-practicalists should prefer the progressiveness approach as well.

I cannot, of course, hope to have convinced everyone by these arguments, nor do I pretend to have said the last word on the subject. But I do hope to have made a strong enough case to at least inspire some to take up the mantle of the progressiveness approach.

Ideal theory has had its day in the sun, but it is time to stop worrying so much about the ideally just society, and to shift our focus to solving problems of injustice and working out how to make our society more progressive. Indeed, if my arguments are correct, then the pursuit of long-term justice may very well depend on us making this shift.

### 5.3. Post-Script: Extensions

This has been an essay about political philosophy. I have asked how political philosophers should theorize about justice, and, in particular, how we should theorize about social reform. But there are various other normative domains in which ideal theory maintains dominance, or at least prominence. Among virtue ethicists it is common to ask what the ideally virtuous agent would be like, or how they would behave. Among theorists of rational choice and epistemologists, it is common to ask similar questions about the ideally rational agent. And consequentialists and contractualists commonly ask what rules the ideal moral code would contain, or what rules would have currency in the ideal moral community.

In each of these domains, it is worth reconsidering whether this focus on ideals is warranted. Should we really be focusing on ideal agents and moral codes, or should we instead be focusing on how agents and codes can get better, and how they can get better at getting better—more progressive? The arguments of this essay suggest that we should at least take seriously the possibility that we should opt for the latter approach. In particular, my arguments against the ideal benchmark view seem perfectly general. If they succeed in showing that we don't need to refer to the ideally just society to make comparative evaluations of more or less just, then it is hard to see why a structurally parallel argument shouldn't also show that we don't need to refer to ideally virtuous or rational agents, or to the ideal moral code, when making comparative evaluations of agents or codes. But it is

admittedly less clear what to think about my arguments against the ideal target view. Insofar as our evaluation of a moral code depends on its consequences, then similar epistemic difficulties would presumably arise when trying to determine how the widespread currency of different moral rules interact to produce outcomes, and so when trying to work out which code is ideal. But it is less obvious whether the same difficulties arise when trying to figure out what the ideally virtuous or rational agent would be like—when trying to work out how different character traits, dispositions, or decision procedures interact to produce more or less virtuous or rational agents.

Even if we do not face precisely the same epistemic difficulties in other domains as we do in political philosophy, however—and even if improving progressiveness therefore takes a very different character in these domains—this is not to say that focusing more on problem solving and progressiveness would not be beneficial. Even anti-practicalists admit the importance of doing practically relevant normative theory; they just deny that practical relevance is all that matters. And work focusing on how to improve, and to get better at improving, our moral codes, our characters, and our rationality, is surely of immense practical import. Indeed, I conjecture that we would be better served by adopting the progressiveness approach not only in political philosophy, but across all normative domains—that doing so would not only help us to better understand justice, social morality, virtue, and rationality, but would also help us both collectively and individually to pursue these normative goals.

For now, however, this remains merely a conjecture. I have defended the progressiveness approach only in the domain of political philosophy. Its full defense must await future work.

## REFERENCES

- Aligica, Paul D. 2018. *Public Entrepreneurship, Citizenship, and Self-Governance*. New York: Cambridge University Press.
- Aligica, Paul D. and Vlad Tarko. 2011. Polycentricity: From Polanyi to Ostrom, and Beyond. *Governance* (2): 237-262.
- Anderson, Elizabeth. 2012. Epistemic Injustice as a Virtue of Social Institutions. *Social Epistemology* 26(2): 163-173.
- Anderson, Elizabeth. 2006. The Epistemology of Democracy. *Episteme* 3 (1-2): 8-22.
- Anderson, Elizabeth. 2010. *The Imperative of Integration*. Princeton: Princeton University Press.
- Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*. Princeton: Princeton University Press.
- Annas, Julia. 1993. *The Morality of Happiness*. New York: Oxford University Press.
- Appiah, Kwame Anthony. 2017. *As if: Idealization and Ideals*. Cambridge: Harvard University Press.
- Aristotle. 2000. *Nicomachean Ethics*. Trans. Roger Crisp. Cambridge: Cambridge University Press.
- Arthur, W. Brian. 2015. *Complexity and the Economy*. New York: Oxford University Press.
- Axelrod, Robert and Michael D. Cohen. 2000. *Harnessing Complexity: Organizational Implications of a Scientific Frontier*. New York: Basic Books.
- Banerjee V. Abhijit and Esther Duflo. 2012. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.
- Barrett, Jacob. 2020a. Comparative Justice as Fairness. Unpublished Manuscript.
- Barrett, Jacob. 2020b. Efficient Inequalities. *Journal of Political Philosophy* 28(2): 181-198.
- Barrett, Jacob. 2020c. Is Maximin Egalitarian? *Synthese* 197: 817-837.

- Barrett, Jacob and Gerald Gaus. forthcoming. Laws, Norms, and Public Justification: The Limits of Law as an Instrument of Reform. In S. A. Langvatn, W. Sadurski, & M. Kumm (eds.), *Public Reason and Courts*. New York: Cambridge University Press.
- Bartels, Larry M. 2017. *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton: Princeton University Press.
- Buchanan, Allen. 2004. *Justice, Legitimacy, and Self-Determination: Moral Foundations for International Law*. New York: Oxford University Press.
- Buchanan, Allen and Robert O. Keohane. 2006. The Legitimacy of Global Governance Institutions. *Ethics & International Affairs* 20(4): 405-437.
- Buchanan, Allen and Russell Powell. 2018. *The Evolution of Moral Progress: A Biocultural Theory*. New York: Oxford University Press.
- Campbell, Donald T. 1979. Assessing the Impact of Planned Social Change. *Evaluation and Program Planning* 2(1): 67-90.
- Campbell, Donald T. 1969. Reforms as Experiments. *American Psychologist* 24(4): 409-429.
- Campbell, Donald T. 1973. The Social Scientist as Methodological Servant of the Experimenting Society. *Policy Studies Journal* 2(1): 72-75.
- Carens, Joseph H. 1981. *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory*. Chicago: University of Chicago Press.
- Carens, Joseph H. 2015. *The Ethics of Immigration*. New York: Oxford University Press.
- Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing it Better*. New York: Oxford University Press.
- Christiano, Tom. 2019. Democracy, Participation, and Information: Complementarity Between Political and Economic Institutions. *San Diego Law Review* 56(4): 935-960.

- Christiano, Tom and Will Braynen. 2008. Inequality, Injustice and Leveling Down. *Ratio* XXI(4): 392-420.
- Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge: Harvard University Press
- Colanders, David and Ronald Kupers. 2016. *Complexity and the Art of Public Policy: Solving Society's Problems from the Bottom Up*. Princeton: Princeton University Press.
- D'Agostino, Fred. 2009a. From the Organization to the Division of Cognitive Labor. *Politics, Philosophy & Economics* 8(1): 101-129.
- D'Agostino, Fred. 2018. How Can We Do Political Philosophy? *Cosmos + Taxis* 5(2): 29-37.
- D'Agostino, Fred. 2009b. *Naturalizing Epistemology: Thomas Kuhn and the 'Essential Tension'*. New York: Palgrave MacMillan.
- Dewey, John. 1922. *Human Nature and Conduct*. New York: Henry Holt and Company.
- Dewey, John. 1927. *The Public and its Problems*. New York: Henry Holt and Company.
- Estlund, David. 2011a. Human Nature and the Limits (If Any) of Political Philosophy. *Philosophy & Public Affairs* 39(3): 207-235.
- Estlund, David. 2011b. What Good Is It? Unrealistic Political Theory and the Value of Intellectual Work. *Analyse & Kritik* 32(2): 395-416.
- Estlund, David 2014. Utopophobia. *Philosophy & Public Affairs* 42(2): 207-235.
- Forstenzer, Joshua. 2019. *Deweyan Experimentalism and the Problem of Method in Political Philosophy*. New York: Routledge.
- Frazer, Michael L. 2016. Utopophobia as a Vocation: The Professional Ethics of Ideal and Nonideal Political Theory. *Social Philosophy & Policy* 33(1-2): 175-192.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.

- Gaus, Gerald. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. New York: Cambridge University Press.
- Gaus, Gerald. (2018). Political Philosophy as the Study of Complex Normative Systems. *Cosmos+Taxis* 5(2): 62-76.
- Gaus, Gerald. 2016. *Tyranny of the Ideal: Justice in a Diverse Society*. Princeton: Princeton University Press.
- Gaus, Gerald. 2019. What Might Democratic Self-Governance in a Complex Social World Look Like? *San Diego Law Review* 56(4): 967-1012.
- Gilbert, Pablo. 2012. Comparative Assessments of Justice, Political Feasibility, and Ideal Theory. *Ethical Theory and Moral Practice* 15(1): 39-56.
- Gilbert, Pablo and Holly Lawford-Smith. 2012. Political Feasibility: A Conceptual Exploration,” *Political Studies* 60(4): 809-825.
- Gladwell, Malcolm. 2000. *The Tipping Point: How Little Things can Make a Difference*. New York: Little, Brown and Company.
- Goodhart, C. A. E. 1984. *Monetary Theory and Practice: The UK Experience*. London: Palgrave Macmillan.
- Goodin, Robert E. 1995. Political Ideals and Political Practice. *British Journal of Political Science* 25(1): 37-56.
- Hamlin, Alan and Zofia Stemplowska. 2012. Theory, Ideal Theory, and the Theory of Ideals. *Political Studies Review* 10(1): 48-62.
- Hayek, F.A. 1960. *The Constitution of Liberty*. London: Routledge and Kegan Paul.
- Hayek, F.A. 2007. *The Road to Serfdom*. Chicago: University of Chicago Press.

- Hayek, F.A. 1964. The Theory of Complex Phenomena. In M. Bunge (ed.) *The Critical Approach to Science and Philosophy: Essays in Honor of Karl R. Popper*. London: Free Press.
- Holland, John. H. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge: MIT Press.
- Holland, John H. 2015. *A Crude Look at the Whole: The Science of Complex Systems in Business, Life, and Society*. New York: Basic Books.
- Kauffman, Stuart A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press.
- Keynes, John Maynard. 2018. *The General Theory of Employment, Interest, and Money*. London: Palgrave MacMillan
- Klein, Ezra. 2020. *Why We're Polarized*. New York: Simon & Schuster.
- Kitcher, Philip. 1993. *The Advancement of Science*. New York: Oxford University Press.
- Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin Company.
- Knight, Jack and James Johnson. 2011. *The Priority of Democracy: Political Consequences of Pragmatism*. Princeton: Princeton University Press.
- Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Landemore, Hélène. 2012. *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: Princeton University Press.
- Lawford-Smith, Holly. 2013. The Feasibility Condition in Political Theory. *Journal of Political Philosophy* 21(3): 27-40.
- Lewontin, Richard. 1978. Adaptation. *Scientific American* 239(3): 157-169.

- Lipsey, R.G. and Lancaster K.J. 1956. The General Theory of the Second Best," *The Review of Economic Studies* 24 (1): 11-32.
- Mahoney, James. 2000. Path Dependence in Historical Sociology. *Theory and Society* 29(4): 507-548.
- March, James G. 1991. Exploration and Exploitation in Organization Learning. *Organizational Science* 2(1): 71-87.
- Martí, José Luis and Philip Pettit. 2010. *A Political Philosophy in Public Life: Civic Republicanism in Zapatero's Spain*. Princeton: Princeton University Press.
- Merton, Robert K. 1936. The Unanticipated Consequences of Purposive Social Action. *American Sociological Review* 1(6): 894-904.
- Miller, John H. and Scott E. Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, Princeton University Press.
- Mill, John Stuart. 1974. *On The Logic of the Moral Sciences*. In J. M. Robson (ed.) *The Collected Works of John Stuart Mill*, volume 8. Toronto: University of Toronto Press.
- Mill, John Stuart. 1977. *On Liberty*. In J. M. Robson (ed.) *The Collected Works of John Stuart Mill*, volume 18. Toronto: University of Toronto Press.
- Mills, Charles W. 2005. 'Ideal Theory' as Ideology. *Hypatia* 20(3): 165-184.
- Mitchell, Melanie. 2009. *Complexity: A Guided Tour*. New York: Oxford University Press.
- Muldoon, Ryan. 2016. *Social Contract Theory for a Diverse World: Beyond Tolerance*. New York: Routledge.
- Müller, Julian F. 2019. *Political Pluralism, Disagreement, and Justice: The Case for Polycentric Democracy*. New York: Routledge.

- North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- North, Douglass C. 2005. *Understanding the Process of Economic Change*. Princeton: Princeton University Press.
- O'Connor, Cailin and James Owen Weatherall. 2019. *The Misinformation Age: How False Beliefs Spread*. New Haven: Yale University Press.
- OECD. 2017. *Debate the Issues: Complexity and Policy Making*. Paris: OECD Publishing.
- Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.
- Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.
- Page, Scott E. 2011. *Diversity and Complexity*. Princeton: Princeton University Press.
- Page, Scott E. 2018. The Imperative of Complexity. *Cosmos+Taxis* 5(2): 4-12.
- Page, Scott E. 2006. Path Dependence. *Quarterly Journal of Economic Studies* 1(1): 87-115.
- Pettit, Philip. 2012. *On The People's Terms: A Republican Theory and Model of Democracy*. New York: Cambridge University Press.
- Platteau, Jean-Philippe. 2000. *Institutions, Social Norms, and Economic Development*. New York: Routledge.
- Popper, Karl. 2013. *The Open Society and its Enemies*. Princeton: Princeton University Press.
- Räikkä, Juha. 1998. 'The Feasibility Condition in Political Theory,' *Journal of Political Philosophy* 6(1): 27-40.
- Rochon, Thomas R. 1992. *Culture Moves: Ideas, Activism, and Changing Values*. Princeton: Princeton University Press.
- Rawls, John. 1999a. *A Theory of Justice: Revised Edition*. Cambridge: Belknap Press.

- Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge: Belknap Press.
- Rawls, John. 1999b. *The Law of Peoples*. Cambridge: Harvard University Press.
- Rawls, John. 2005. *Political Liberalism: Expanded Edition*. New York: Columbia University Press.
- Riker, William H. 1982. *Liberalism Against Populism: A Confrontation Between The Theory of Democracy and the Theory of Social Choice*. San Francisco: W.H. Freeman and Co.
- Robeyns, Ingrid. 2008. Ideal Theory in Theory and Practice. *Social Theory and Practice* 34(3): 341-262.
- Rosenberg, Alex. 2016. On The Very Idea of Ideal Theory in Political Philosophy. *Social Philosophy and Policy* 33(1-2): 55-75.
- Sabel, Charles F. and Jonathan Zeitlin. 2012. Experimentalist Governance. In D. Levi Faur (ed.) *The Oxford Handbook of Governance*. New York: Oxford University Press.
- Sabel, Charles F. and Jonathan Zeitlin. 2010. *Experimentalist Governance in the European Union: Toward a New Architecture*. New York: Oxford University Press.
- Schaefer, Alex. 2019. Coping with Complexity: A Theory of Hayekian Interventionism. In D. J. Boudreaux, C. J. Coyne, and B. Herzeberg (eds.), *Interdisciplinary Studies of the Political Order: New Applications of Public Choice Theory*. New York: Rowman & Littlefield.
- Sen, Amartya. 2009. *The Idea of Justice*. Cambridge: Belknap Press.
- Schelling, Thomas. 1971. Dynamic Models of Segregation. *The Journal of Mathematical Sociology* 1(2): 143-18.
- Schelling, Thomas. 1978. *Micromotives and Macrobehavior*. New York: W.W. Norton & Co.
- Schmidtz, David. 2011. Nonideal Theory: What It Is and What It Needs To Be. *Ethics* 121(4): 772-796.

- Schmidtz, David. 2016. A Realistic Political Ideal. *Social Philosophy and Policy* 33 (1-2): 772-796.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Shelby, Tommie. 2016. *Dark Ghettos: Injustice, Dissent, and Reform*. Cambridge: Harvard University Press.
- Shelby, Tommie. 2013. Racial Realities and Corrective Justice: A Reply to Charles Mills. *Critical Philosophy of Race* 1(2): 145-162.
- Sidgwick, Henry. 1981. *The Methods of Ethics*, 7<sup>th</sup> ed. Indianapolis: Hackett.
- Simmons, A. John. 2010. Ideal and Nonideal Theory. *Philosophy & Public Affairs* 38(1): 5-36.
- Simon, Herbert. 1962. The Architecture of Complexity. *Proceedings of the American Philosophical Society* 106 (6): 467-482.
- Sinnott-Armstrong, Walter. 1999. Some Varieties of Particularism. *Metaphilosophy* 30(1/2): 1-12.
- Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Methuen & co.
- Smith, Peter. 1998. *Explaining Chaos*. New York: Cambridge University Press.
- Somin, Ilya. 2010. Foot Voting, Political Ignorance, and Constitutional Design. *Social Philosophy & Policy* 28(1): 202-227.
- Stemplowska, Zofia. 2008. What's Ideal about Ideal Theory? *Social Theory and Practice* 34(3): 319-340.
- Stuntz, William J. 2000. Self-Defeating Crimes. *Virginia Law Review* 86 (8): 1871-1899
- Swift, Adam. 2008. The Value of Philosophy in Nonideal Circumstances. *Social Theory and Practice* 34(3): 363-387.

- Tenner, Edward. 1996. *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*. New York: Knopf.
- Tetlock, Philip E. 2006. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press.
- Tetlock, Philip E. and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction*. New York: Crown Publishers.
- Valentini, Laura. (2012). Ideal vs. Non-Ideal Theory: A Conceptual Map. *Philosophy Compass* 7(9): 654-664.
- Valentini, Laura. 2011. A Paradigm Shift in Theorizing about Justice? A Critique of Sen. *Economics & Philosophy* 27(3): 297-315.
- Vallier, Kevin. 2014. *Liberalism and Public Faith: Beyond Separation*. New York: Routledge.
- Wiens, David. 2012. Prescribing Institutions Without Ideal Theory. *The Journal of Political Philosophy* 20(1): 45-70.
- Wiens, David. 2015a. Against Ideal Guidance. *The Journal of Politics* 77(2): 433-445.
- Wiens, David. 2015b. Political Ideals and the Feasibility Frontier. *Economics and Philosophy* 31(3): 447-477.
- Wilson, David S. 2016. Two Meanings of Complex Adaptive Systems in (ed.) D.S. Wilson and A. Kirkman, *Complexity and Evolution*. Cambridge: MIT Press.
- Wolfe, David A. 2018. Experimentalist Governance: Conceptual Approaches and Practical Cases. Background paper for an OECD/EC Workshop on 14 December 2018 within the workshop series “Broadening innovation policy: New insights for regions and cities,” Paris.