Social Reform in a Complex World

Jacob Barrett

1.   Introduction

We live in an unjust world. Our social and political institutions stand in need of reform. But of all the changes we might make to these institutions, which would genuinely promote justice? And how should we, as theorists, go about trying to figure this out?

Perhaps the most straightforward approach is *problem solving*: diagnosing particular problems of injustice in our world, and proposing narrowly targeted institutional solutions. For example, we might aim to identify actual instances of status and resource inequality, of discrimination and oppression, of human rights violations and unjustifiable restrictions on freedom. And we might work to uncover their causes, and to come up with changes to our laws, policies, or social norms that would mitigate or eliminate them. In other words, we might adopt a relatively narrow and short-term perspective, aiming to identify "remediable injustices" in our world along with promising institutional remedies—remedies that would promote justice by chipping away at the many problems of injustice that confront us.[1]

Another is *ideal theory*. On this approach, we begin not with the injustices we currently face, but by attempting to outline what the overall best institutional arrangement would be, before then figuring out "how this long-term goal might be achieved, or worked towards, usually in gradual steps."[2] So instead of diagnosing specific problems of injustice,

---

[1] In describing the problem solving approach I hew most closely to Wiens 2012, but see also Sen 2009—from whom I borrow the term "remediable injustice," vii—Anderson 2010, ch. 1, and Schmidtz 2011.

[2] Rawls 1999a, 89. Rawls labels theorizing about how to make progress towards the ideal "nonideal theory," but I here use the term "ideal theory" to refer to this entire approach to

we aim to identify the most just, problem-free institutional arrangement we could ever achieve. And instead of working out which changes to our current arrangement would mitigate or eliminate present injustice, we ask which changes would constitute progress towards this ideal. Our focus is therefore more comprehensive than on the problem solving approach, concerned with the ideal institutional arrangement as a whole rather than with targeted solutions to particular problems. And it is also longer-term, concerned not with improving justice in the short term but, most centrally, with identifying a long-term goal—an end-goal—"to guide the course of social reform."[3]

When we compare these two approaches side by side, ideal theory might seem alarmingly farsighted. Why would anyone think that we should focus on making progress towards a far-off goal of ideal justice, rather than on diagnosing and solving the problems of injustice we face right now? Isn't the whole point of theorizing about social reform to uncover ways of ameliorating present injustice? But to this charge of hyperopia, ideal theorists retort 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 "retard," "stall," or "permanently block… movement towards

theorizing about reform, which includes both identifying the ideal and working out how to make progress towards it. I also set aside the interpretive question of whether Rawls aimed to identify an ideal institutional arrangement or only principles to govern such an arrangement—my concern, regardless, is with *institutional* ideal theorists who attempt the former. Of course, there are many other ways to use the term "ideal theory" as well, on which, see Hamlin and Stemplowska 2012 and Valentini 2012.

[3] Rawls 1999b, 215.

overall justice."[4] So when evaluating potential reforms, we cannot focus only on their short-term effect on particular problems of injustice, but must balance this against their promotion of greater justice in the long term. And while problem solving may be a good way to figure out the first half this balance, it is not enough. We must also take into account overall long-term justice, and for that, we need ideal theory.

Ideal theorists have a point. Our world is complex—it is composed of many interacting parts—and, as I shall explain, this complexity entails that ameliorating particular problems of injustice may indeed set back the achievement of greater long-term justice. So theorizing about social reform does require more than just problem solving, as ideal theorists rightly argue. But it does not follow that we should supplement problem solving with ideal theory. In fact, I will argue that the very complexity that generates a conflict between ameliorating immediate problems of justice and promoting overall long-term justice also renders ideal theory epistemically overdemanding for beings like us: it makes it impossible for us either to identify the ideal or to track our progress towards it, at least with sufficient confidence to warrant its pursuit. So, thanks to complexity, problem solving is unsatisfactory and ideal 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.

The answer I propose is that rather than attempting to identify ways of making

---

[4] Simmons 2010, 21. Compare Buchanan 2004, ch. 1, Robeyns 2018, and Valentini 2012. Some ideal theorists go further, arguing that we cannot even address "pressing problems" of injustice without a conception of the ideal in hand (Rawls 1999b, 8). But this view has been subject to such a deluge of recent criticism by Anderson 2010, Sen 2009, Wiens 2015a and others that it has been all but abandoned. I therefore set it aside here.

progress towards the ideal, we should instead approach questions of long-term justice by working out how to make our institutional 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 institutional goal. And, more concretely, I argue that the progressiveness of an institutional arrangement depends on its ability to flexibly implement many promising solutions to problems as they arise, to select for those solutions that prove successful while eliminating those that do not, and to help us learn from both our successes and our inevitable failures. On this approach, problem solving has a place, but the solutions it generates are viewed as hypotheses that function within a broader framework of institutional experimentation, selection, and learning.[5] But ideal theory has little (if any)[6] place: theorizing about overall long-term justice instead takes the form of figuring out how to enhance the progressiveness of this overarching framework.

---

[5] As we will see, some problem solvers endorse a similarly limited role for problem solving, and I would be attacking a straw man if I criticized them while claiming otherwise. But my purpose in this paper is not to criticize problem solving per se. It is to investigate the limits of problem solving and explore how we might go beyond them.

[6] I say "if any" because my criticism of ideal theory concerns only its ability to provide us with a long-term goal for reform, not its relevance or value in general. For all I say here, ideal theory might, for example, help us to appreciate how existing arrangements fall short of ideal justice or to uncover our basic evaluative criteria. See Swift 2008 and Gilabert 2012. And ideal theory might also be valuable in its own right, independently of its contribution to theorizing about social reform (which is my only concern here). See Estlund 2014. Thanks to an anonymous reviewer for urging me to say more about this point.

2. Complexity and Problem Solving

Let us begin with some terminology. An "institutional arrangement" as I use the term, refers to a set of formal and informal institutions (for example, laws and social norms), as well as the background conditions that are causally relevant to their functioning (for example, facts about the natural environment, demographics, and technology). Both institutions and background conditions are what I call "institutional features": they serve as the inputs to our (often implicit) models of how an institutional arrangement produces its effects. The features represented by the outputs of such models I call "outcomes," and by the "justice" of an institutional arrangement I mean an evaluation of the outcome it produces given some criterion of justice. This may be an external criterion that takes into account and balances such factors as freedom, equality, oppression, and procedural justice.[7] 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 evaluative criteria of those living under it, rather than against independently specified values or principles. Throughout, I make no attempt to defend any particular criterion of justice, but instead invite the reader to apply her favored criterion to the issues at hand. I assume only that one is not an "institutional fundamentalist" who denies that the outcome produced by an institutional arrangement is at all relevant to its justice.[8]

I emphasize the distinction between an institutional arrangement, the outcome it

[7] Abstract values such as freedom and equality are sometimes themselves referred to as "ideals." This is a perfectly fine use of language, but, to avoid ambiguity, I will never employ it myself, and will instead reserve the term "ideal" as a shorthand for "the ideally just institutional arrangement." Thanks to an anonymous reviewer for raising this issue.

[8] For a critique of such fundamentalism, see Sen 2009, ch. 3.

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 an institutional arrangement's justice is always (logically, though perhaps not temporally) a two-step procedure.[9] We must first map an institutional 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 the distribution of income? And the second is to evaluate that outcome given our chosen criterion: would the predicted change in such variables amount to a net increase or decrease of justice, understood, say, in Rawlsian, utilitarian, or libertarian terms? Though perhaps the bulk of political philosophy concerns the appropriate criterion to use at this second step, my focus here is the first step where we map institutional arrangements to outcomes. For convenience, I will therefore often speak as if we can skip the second step and map institutional arrangements to justice directly. But it is important to remember what this involves. It requires us to employ a "predictive model" of how various institutional features interact to produce an outcome, which we must then evaluate in terms of its justice.[10]

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.[11] In analyses of societies as complex systems, these "parts" are

[9] Compare Gaus 2016, ch. 2.

[10] Ibid.

[11] On complexity in general, see Mitchell 2009, and for examples of its recent application to economics, see Arthur 2014, to public policy, see Colanders and Kupers 2016, and to political philosophy, see Gaus 2016, 2018.

usually thought of as people, whose interactions produce emergent patterns that no individual intended or perhaps even foresaw.[12] The classic example of this is Smith's discussion of how, given certain institutional arrangements, the market interactions of individuals each pursuing their own self-interest leads to greater social welfare.[13] 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.[14] And 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.[15] 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).

An understanding of the various ways that individuals may unwittingly interact to produce both good and bad emergent phenomena in the presence of different institutional arrangements—which, after all, structure such interactions by constraining and incentivizing different forms of behavior—is important to any analysis of social change. In the first place, it dispels us of both the overly rationalistic view that individuals can only produce just outcomes by explicitly aiming to do so, and the overly complacent view that individuals

---

[12] For an overview of this sort of complexity, see Miller and Page 2007.

[13] Smith 1776.

[14] 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.

[15] Schelling 1971. Of course, much actual segregation is produced very intentionally, rather than in this way.

pursuing their own projects reliably promote justice in all circumstances: both views fail to recognize that the causal relation between the achievement of justice and the intentional pursuit of other goals crucially depends on which institutions are in place.[16] In the second, it helps to explain why the effects of institutional change are so difficult to predict, as it is one important source of the unanticipated consequences that often accompany such changes.[17] Going forward, however, my primary concern will be not with the micro-level complexity that characterizes interactions between individuals, but instead with two forms of macro-level or institutional complexity. It will be with the way that different institutional features interact to produce outcomes, as well as to produce changes in other institutional 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 institutional features. As is now commonplace among institutional theorists, the operation of any one institution is importantly dependent on the presence and operation of other institutions, as well as on background conditions.[18] Often this phenomenon is discussed by economists under the rubric of the "general theory of second best." As Lipsey and 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

[16] Compare Wilson 2016.

[17] For a classic analysis of unanticipated consequences, see Merton 1936. For a more recent overview chock full of examples, see Tenner 1996.

[18] See especially North 1990.

might be Pareto superior (better for some and worse for none) than those that fail to meet only one.[19] This, however, is just one instance of the more general phenomenon of combinatorial complexity. If we cannot have the optimal or "ideal" institutional 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 institutional features cannot be reduced to the sum of the effects of each feature by itself. So even when each of two institutional 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 a background of social trust and the absence of norms prohibiting profit seeking.[20] 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 abolish.[21] Color-blind policies may seem just in isolation, but may only serve to reinforce racial inequalities produced by other features of our institutional arrangement; in such cases, color-conscious policies may promote justice, even if they would undermine it given background equality.[22] 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

---

[19] Lipsey and Lancaster 1956-1957.

[20] This point goes back to Smith 1776, but see also Platteau 2000, chs. 5, 7.

[21] Gaus and I discuss this point at length in Barrett and Gaus forthcoming.

[22] Compare Anderson 2010, ch. 8.

programs might improve justice, and opening our borders might improve justice, but doing both together might be disastrous: the influx of immigrants might result in the entitlement programs being stretched beyond the breaking point.[23]

These instances of combinatorial complexity all involve features interacting to produce outcomes. But institutional features also interact in the sense that prior institutional changes may further or set back later institutional changes, and this gives rise to a type of path-dependency that I will call *transitional complexity*. For example, even if opening our borders and enacting more generous entitlement programs would indeed improve justice, opening our borders first might make it more difficult to provide more generous entitlement programs later, if (as some empirical evidence suggests) influxes of immigration undermine public support for such programs.[24] More general phenomena relevant to transitional complexity include lock-in and backlash. Lock-in occurs when a change prevents further changes, often because it generates interest groups who are able to maintain the new status quo.[25] Backlash occurs when an institutional 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.[26] Perhaps even more obviously, changes to secondary rules (rules for changing other rules) clearly interact with future changes to our institutions. For example, in a democracy, different voting

---

[23] Related worries are common in the literature on immigration. See Carens 2013 ch. 12.

[24] Ibid.

[25] See North 1990, 94-99.

[26] 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.

schemes predictably lead to different policies, as do different constitutional constraints on what is subject to democratic rule.[27]

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 particular institutional features, and identify institutional 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 health care 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 sort of economic, health care, and other social safety net programs are in place, whose cost and efficacy depends on education policies, which also affect 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 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

---

[27] Riker 1982 provides an authoritative discussion of the way that different voting rules may produce different results even given the same voter preferences.

which further changes are likely to be democratically authorized. (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 we cannot focus only on identifying institutional 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 institutional features.[28] And this suggests that we must expand our vision along two dimensions. First, we must adopt a more comprehensive, holistic attitude to evaluating institutional 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 reform. Taking this suggestion to its limit, we might adopt an ideal-theoretic orientation, first trying to identify an ideally just institutional arrangement to serve as a "long-term goal of political endeavor," and then attempting to work out how to make progress towards it.[29] This approach, after all, seems tailor-made to accommodating complexity: to identify an institutional arrangement as

[28] Simmons 2010, 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.

[29] Rawls 1999b, 138.

ideal, we must take into account the combinatorial interactions of all its component features, and to determine whether a short-term change constitutes progress towards this ideal, we must take into account all relevant instances of transitional complexity. But, as we will now 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. Epistemic Asymmetries and Ideal theory

The view that we ought to set the ideal as our long-term target for reform must be distinguished from the closely related view that to reform our institutional arrangements is to make them better approximate or resemble the ideal. Indeed, an appreciation of combinatorial and transitional complexity undermines whatever initial plausibility this latter view might have. If there were no combinatorial complexity, then the justice of an institutional arrangement would depend on its similarity to the ideal: every time we implemented an institutional feature that obtains at the ideal, this would make our institutional arrangement more just. If there were no transitional complexity, then progress towards the ideal would depend on similarity to the ideal: every time we implemented an institutional feature that obtains at the ideal, we would make progress towards the ideal, since there we would be less changes left to make. But since both sorts of complexity exist, neither relation holds. Making institutional arrangements more similar to the ideal can, due to combinatorial complexity, result in a decrease in justice, or, due to transitional complexity, constitute progress away from the ideal.[30] And this implies that making our institutional

---

[30] 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 certainly aware of transitional complexity, the model they employ assumes that

arrangement better approximate or resemble the ideal is not a reliable way to promote justice. If ideal theory is to stand any chance of being a viable approach to theorizing about social reform, its goal cannot merely be to identify ways of making our current arrangements more similar to the ideal. Instead, it must be to identify "steps" that constitute progress towards the ideal, where making progress towards the ideal is understood in such a way that it sometimes involves making our arrangement less similar to the ideal along the way.[31]

Ultimately, I will argue that ideal theory is not, in fact, a viable approach to theorizing about social reform. But before we get to this critique, we need a better picture of ideal theory in mind. This requires us to answer three questions. First, what do we mean by the "ideal" institutional arrangement? Second, how should we cash out the idea of making "progress towards" the ideal if—as we have just seen—it does not merely amount to making our arrangement better resemble it? And third, what role is ideal theory supposed to play in a full account of social reform?

The first answer is straightforward enough. To say that something is ideal is to say that it is best, and, in this context, "best" means "most just." But in order for the ideal to serve as a long-term goal, it is not enough for it to be the most just institutional arrangement that is (say) conceptually possible—it must be possible in the sense that there is some feasible path between it and us. Of course, this idea is very rough. The feasibility of an institutional arrangement depends both on what our world 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. Still, no matter the details, the key point

---

making an institutional arrangement more similar to another constitutes progress towards the latter. Thanks to transitional complexity, this is not always the case.

[31] Simmons 2010, 23.

is that our judgment about whether something is feasible depends on a prediction of whether we can get there from here. To identify the ideal, we cannot simply form a conception of what a *perfectly* or *fully* just arrangement would be like given our criterion of justice, since such an arrangement might very well be infeasible, and therefore unable to play the role of a long-term goal. Instead, we must form a prediction of which institutional arrangements we could eventually realize, and then another prediction of which, of these, would produce the most just outcome—as always, given our chosen criterion.[32]

Turn, then, to the second question: how should we understand "progress towards" the ideal? Well, consider the earlier worry that implementing solutions to particular problems of injustice might "retard," "stall," or "permanently block… movement towards overall justice."[33] This worry suggests that progress should be understood in temporal terms: we make progress towards 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 towards the ideal is to increase the probability we will eventually get there.[34] This introduces a conflict internal to the notion of "progress towards": for example, a step down a revolutionary path may have less chance of taking us

[32] Compare Buchanan 2004, 61-63 and Wiens 2015b.

[33] Simmons 2010, 21.

[34] 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 Gilabert and Lawford-Smith 2012 and Lawford-Smith 2013.

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 towards")[35] into a single standard for judging how "far" the ideal is from a particular arrangement. In determining whether an institutional change would constitute progress towards the ideal, we must therefore form a prediction of how the resulting arrangement will continue to change. We must forecast forward from this initial change, asking what effect it will have on the probability we will ever reach the ideal, the time it will take us to do so, and the extent to which it therefore constitutes progress towards or away from it.

Finally, what role does ideal theory in a full account of social reform? Now, if identifying and pursuing the ideal is worth doing, it must be because there are some cases where implementing short-term improvements sets back progress towards the ideal. Otherwise, there would be no need to identify an ideal in order to make progress towards it: every short-term improvement would simultaneously constitute progress towards the ideal, so we could make such progress simply by implementing the short-term improvements we discover through problem solving.[36] But, as we have seen, complexity does generate a tradeoff between short-term justice and progress towards 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 towards the ideal, and we should not saddle ideal theorists with such an extreme commitment: as they emphasize,

[35] For example, we might also factor in the morally relevant cost of achieving the ideal. See Räikkä 1998. But I prefer to analyze such costs separately, as something to be traded off against making progress to the ideal rather than as an element of such progress.

[36] Gaus 2016, 82-84.

there are times when ignoring short-term injustice in order to make progress towards the ideal is morally impermissible or otherwise not worth the cost.[37] So we should understand ideal 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, at least in 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.

So understood, ideal theory is both maximally comprehensive and maximally long-term, and it might therefore seem to fully accommodate the complexity of our world. But, alas, the very complexity that makes ideal theory attractive also makes it impracticable—at least for agents like us. For combinatorial and transitional complexity not only ensure that short-term solutions to particular problems of injustice sometimes conflict with progress towards 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 institutional arrangement, our predictions about what outcomes those changes will produce, and therefore about what their effects on justice will be, decrease in reliability. Second, due to transitional complexity, as we forecast the effects of institutional changes further into the future, our predictions about which subsequent changes will occur become less reliable as well. And these two asymmetries undermine the epistemic presumptions of ideal theory. We cannot identify the ideal institutional 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 towards it with this requisite degree of confidence.

Consider first a recent argument of Gaus's, which begins by noting that our

---

[37] Rawls 1999a, 89; Simmons 2010, 18-22.

predictive models of how institutional arrangements interact to produce outcomes are not very 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[38]—and, furthermore, that like all models of complex systems, they are subject to "error inflation."[39] We may calibrate our models of actual institutional arrangement to the data: if we predict, say, that increasing the minimum wage will spike unemployment, or that lowering the corporate tax rate will increase inequality, but find that this doesn't occur—or that either change produces some wholly unanticipated effect, say, on the gendered or racial division of labor—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 institutional arrangements that differ more and more in their institutional 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."[40]

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 the corporate tax rate, than in our prediction of a change to both, since the errors we make in predicting the

---

[38] These two possibilities track Knight's 1921 distinction between probabilistic "risk" and non-probabilistic "uncertainty."

[39] Gaus 2016, 80. Compare Smith 1998.

[40] Gaus 2016, 80.

effect of each carry over into our prediction of how they interact. 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, since such systems differ from actual ones in so many ways that errors massively inflate. And we should have even less confidence—indeed, basically none at all—in our ability to predict the outcome produced by an even more divergent arrangement, designed to handle not only economic injustice, but racial and gender injustice, global injustice, and all other forms of injustice as well. So, since most candidate ideals differ radically from any actual institutional arrangement, we lack the ability to make a confident prediction about the outcome such arrangements will produce, and, therefore, to evaluate their justice. An essential presupposition of ideal theory cannot be met: we cannot judge which institutional arrangement is ideal, and therefore worth pursuing at the expense of short-term improvements, with sufficient confidence to license this pursuit.

Gaus's epistemic critique of ideal 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. To begin, note that, 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, and so on. And this general tendency is magnified in complex systems, where, due to transitional complexity, it becomes impossible for us to predict anything in the very long-term. For, 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

institutional 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 institutional changes requires us to predict what outcome those changes will realize. 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 institutional 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 and 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 institutional change far into the future at all: we cannot determine which institutional arrangements we could get to in the very long-term, nor which short-term changes bring us towards them, at least with any reasonable degree of confidence.[41]

On the basis of similar considerations, 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."[42] In other words, whereas Gaus worries that we cannot figure out how just various candidate ideals would be,

---

[41] Compare Hayek 1960, ch. 2 and Tetlock 2006.

[42] Wiens 2015b, 467.

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."[43] And since, as I have emphasized, identifying the ideal requires us to figure out both of these things—to determine which *feasible* institutional arrangement is *most just*—this suggests that ideal theory is an impossible enterprise.[44] Indeed, the problem is even worse than Gaus or Wiens suggests, since even if we somehow knew which institutional arrangement was ideal, we would still lack the ability to confidently identify which short-term changes would constitute progress towards 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 institutional changes far into the future this way. 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 outcomes will be produces—something we cannot do when it comes to institutional 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.

Summing up our discussion, then, ideal theory requires us to confidently determine

---

[43] Ibid.

[44] As Buchanan 2004, 61-63, argues, if a conception of the ideal is to serve as an appropriate long-term goal for reform, it must be not only *causally* feasible but also *morally* feasible in the sense that it can be reached through morally permissible means. This only adds to the difficulty of determining which institutional arrangements are feasible, and, therefore, which is ideal.

three things: which institutional arrangements are feasible, which of these is most just, and which short-term changes constitute progress towards this ideal. But, thanks to complexity, we cannot perform any of these tasks—let alone all three—so we cannot identify steps that would constitute progress towards the ideal with sufficient confidence to warrant taking these steps at the expense of short-term justice. Though focusing only on identifying narrow solutions to particular problems of injustice is not a reliable way to promote overall long-term justice, trying instead to figure out how to make progress towards the ideal is beyond our epistemic ken. The trick will be to see if we can thread the needle, and avoid both the myopia of problem solving and the epistemic overdemandingness of ideal theory.[45]

4.   Experimentation and Progressiveness

So far, we have seen that, in theorizing about social reform, we face two competing pressures. Because of the potential for interaction between different narrow solutions to immediate problems of injustice, there is a pressure to expand our sights outwards and forwards towards more comprehensive changes and their longer-term effects. But at the same time, our ability to predict the effect of institutional change rapidly deteriorates as we attempt to expand our sight in either of these ways. So there is a contravening pressure to contract our focus back to the narrow and the short-term.

---

[45] One further worry for ideal theorists is that there may be no fixed ideal institutional arrangement to pursue in the first place, since, due to combinatorial complexity, "institutions adopted for a particular time, even if optimal... at that time, may be far from optimal as the human environment changes over time" (North 2005, 132). I lack the space to develop this "moving target" objection here, but see Muldoon 2016, 6, 29, Popper 2013, and Rosenberg 2016, 64-70. Thanks to an anonymous reviewer for suggesting I flag this worry.

Now, to be clear, the problem raised by these competing pressures is not just one for ideal theorists, nor 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 towards, say, medium-sized, medium-term changes, there would remain a gap between the largest and longest-term reforms whose effects we could confidently identify and pursue, and more comprehensive institutional changes that would produce greater justice in the longer term. And implementing these medium-sized, medium-term changes would still risk setting back the pursuit of overall longer-term justice, either by interacting to undermine overall justice (due to combinatorial complexity), or to set back future improvements (due to transitional complexity), just like narrow, short-term changes. This leaves us with our central methodological challenge: *how can we identify institutional reforms that we can predict with reasonable confidence will promote overall long-term justice, when we cannot figure out which institutional arrangements to aim at in the long-term, nor how to make progress towards them?*

Although problem solvers rarely address this question directly, they sometimes appear to suggest 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 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."[46] But since we are not very good at this, we must recognize that the solutions we generate are "tentative and experimental.[47] We must give up on anything approaching certainty and, as Anderson puts it, treat "imagined solutions to identified problems… as *hypotheses*, to be tested in

[46] Wiens 2012, 66.

[47] Ibid.

experience."[48]

Problem solvers who endorse this experimental orientation are on the right track, and we are now well positioned to understand why. In theorizing about social reform, we face two epistemic asymmetries: we are worse at predicting the effects of larger changes than smaller changes, and at predicting the longer-term effects of changes than their shorter-term effects. Indeed, at the limit, we are much worse at predicting the effect of any change than we are at evaluating our current arrangement, since every institutional 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 institutional 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 is why the ultimate 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 institutional reforms that are "worth a try,"[49] but whose actual effects on justice can only be ascertained through trying them out in various combinations—that is, through social experimentation.[50]

Once we recognize our epistemic limitations and the resultant need to test our proposals in practice, however, this mandates a far more significant shift in our approach to theorizing about social reform than problem solvers seem to realize. To see why, let us

[48] Anderson 2006, 6.

[49] Schmidtz 2015, 772.

[50] 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 1978, Dewey 1927, and Popper 2013.

consider the question of what makes a proposed solution worth a try. One factor is the effect we predict it will have on problems of injustice, taking into account complexity to whatever extent we can. But that is not all. For one thing, it also matters how much experiments teach us about how to improve justice in the future, since some experiments provide more useful data than others that we may feed back into our causal models to develop more promising institutional solutions going forward. Most obviously, we learn more from novel institutional experiments than from those we are already familiar with: implementing such solutions allows us to generate new data by *exploring* the space of institutional possibilities rather than merely *exploiting* our current knowledge of what has worked in the past.[51] Perhaps less obviously, we must also take into account the epistemic quality of the experiments we engage in. For example, more radical experiments generally have less internal validity than more modest ones, since their effects depend on the interactions between so many variables that it is difficult to trace out the causation. But there is also the worry that small-scale social experiments lack external validity: that their effects will not "scale up" to the societal level.[52] Thankfully, these are not mutually exclusive alternatives: we may also engage in relatively modest experiments at relatively large scales, arguably maintaining a reasonable level of internal and external validity.[53] Of course, there is much room for further debate here concerning the epistemic merits of different types of social experiments, and there is already a large literature on the subject.[54] But the general

[51] March 1991 provides a classic discussion of the exploration/exploitation tradeoff. See also D'Agostino 2009.

[52] Gaus 2016, 89-93.

[53] Popper 2013, 176.

[54] For a start, see Shadish, Cook, and Campbell 2002 and Cartwright and Hardie 2012.

point is that, in thinking about which experiments are worth a try, we must consider not only their predicted effects on justice, but also what we expect to learn from them. Sometimes, the better experiment may not be the one that we predict to be more just in the short-term, but the one that will teach us more going forward.

This might seem more like mad science than social reform. Why should we forgo ameliorating present injustice in order to gather social scientific data? The answer is that we face a tradeoff between short-term and long-term justice, and that the better models we have of how institutional features interact to produce outcomes, the better positioned we are to promote justice in the future. Of course, we should not always forgo short-term improvements for this reason, maniacally expanding our knowledge 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,[55] or because their costs systematically fall on already disadvantaged groups.[56] But just as ideal theorists argue that we must sometimes forego short-term justice in order to make progress towards the ideal, my suggestion is that we must sometimes forego short-term justice in order to better position ourselves to promote further justice. We must trade off the predicted short-term justice produced by a social change not against how much it constitutes *progress towards* 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 goal. And one factor that is relevant to the progressiveness of an institutional arrangement is how well we understand how to improve justice from there.

[55] Müller 2019, 176-177.

[56] Knight and Johnson 2011, 48-49.

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 institutional features interact is one way for us to improve our progressiveness, we can similarly improve our 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. But at the institutional level, we might also improve the social epistemic conditions in which we theorize about reform. To some extent, this depends on the existence of social epistemic conditions that are conducive to good inquiry in general, including free speech, a diversity of research agendas, shared vocabularies, and so on, that philosophers of science, at least since Kuhn, have made much progress exploring.[57] But it also depends more specifically on the extent to which we have mechanisms in place for monitoring the institutional experiments we engage in, gathering information from the experiences of other past and present societies, and storing it in our institutional memory.

Putting this idea more generally: just as we earlier understood making *progress towards* 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 an institutional arrangement as increasing the probability and speed with which it will continue to improve in justice indefinitely into the future. But the progressiveness of our institutional arrangement depends not only on its amenability to learning; it also depends on the extent to which our arrangement permits flexible experimentation going forward. Among other

[57] Kuhn 1962. See also Kitcher 1993, D'Agostino 2009, and, for a discussion of social epistemic conditions that are conducive to good *moral* inquiry in particular, Buchanan and Powell 2018.

things, this requires the replacement of norms of dogmatism with those permitting experimentation,[58] the ability to avoid lock-in due to seizure by interest groups,[59] and a reluctance to implement policies that are difficult to reverse.[60] But since our goal is to promote justice, it would be too simplistic to think that we always ought to avoid lock-in, that we always want to leave all options open going forward. Sometimes, we do want to close options off, at least temporarily: we want institutions that genuinely solve problems of injustice to remain stable as long as they remain solutions. And this mandates a reliance not only on learning mechanisms that help us to update our causal models in response to feedback about the effect of our experiments, but also on selection mechanisms by which we can stabilize successful experiments and eliminate failed ones.

To unpack this idea further, let us say that an institutional 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 "worth a try." And let us say that an institutional experiment is a "success" if it is worth keeping, a "failure" if it is not. Now, the reason we have to engage in institutional experimentation in the first place is that we cannot confidently predict which institutional 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 institutional features we implement will prove worth keeping, nor how long they will remain that way—especially as we attempt to forecast the effects of these changes further into the future. What we need,

---

[58] This is a running theme of both Dewey 1927 and Popper 2013.

[59] North 2005, 125.

[60] Campbell 1969, 410.

then, are selection mechanisms that allow us to stabilize institutional 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 an institutional framework that embodies mechanisms of epistemic and practical feedback—of learning and selection—is the only way that agents like us can promote long-term justice. Due to combinatorial complexity and the epistemic asymmetry to which it gives rise, we cannot hope to predict the effects of large-scale long-term changes (or the sequential combinations of smaller changes). So the only way for us to determine the effect of combinations of institutional features is to try them out and monitor them after the fact. Due to transitional complexity and the epistemic asymmetry it produces, we cannot hope to predict what the effect of short-term changes will be on the institutional arrangement we will end up with in the long run. So the only way for us to pursue long-term reform is through mechanisms that allow for continual adjustments of our institutional arrangement on the basis of the feedback we gather from this experimentation. Thus, whereas experimentation and learning mechanisms of epistemic feedback allow us to tame the epistemic difficulties raised by combinatorial complexity by expanding our predictive capacities and reducing the error that goes into such predictions, the goal of selection mechanisms of practical feedback is not to tame but to harness transitional complexity by reducing our reliance on prediction through the correction of errors after the fact. 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 institutional 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.

This is all rather abstract. To make it more concrete, let us examine two approaches to realizing these mechanisms currently popular in the literature. The first is experimental democracy.[61] Here, the rough idea is that experimentation is achieved through democratic deliberation and voting on which reforms to implement, and selection and learning occur through the monitoring of these institutions' effects and then deliberating and voting on which institutions to maintain and which to eliminate. Thus, experimentation is achieved primarily through democratically authorized "reforms as experiments,"[62] selection primarily through deliberating and voting on which experiments to maintain, and we learn primarily by gathering evidence about the effects of these various reforms. Long-term 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.[63]

This brings us to the second major approach: polycentricity.[64] This time, the rough idea is that experimentation is achieved not only through consecutive or diachronic

[61] For historical defenses of experimental democracy, see Dewey 1927 and Popper 2013. For more recent defenses, see Anderson 2006 and Knight and Johnson 2011.

[62] Campbell 1969.

[63] Knight and Johnson 2011, ch. 6.

[64] Here, we find an approach primarily advocated for by political economists such as Ostrom 2005 and Aligica and Tarko 2011, but also more recently by philosopher such as Müller and Gaus 2016, ch. 4.

experimentation, but also through a number of institutions being tried out simultaneously in different jurisdictions. 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.[65] In each case, experimentation is achieved through different groups employing different laws, policies, or norms at the same time; selection occurs as groups or rules compete for adherents; and learning arises through groups monitoring the results not only of their own (formal or informal) institutions, but also those of others, and adjusting their own institutions accordingly.

We need not settle here the debate between experimental democrats and polycentrists. Indeed, the conflict between the two approaches is less stark than I have just made it seem. Experimental democrats, for example, generally recognize the importance of some degree of polycentricity, as well as the necessity of not only formal governmental procedures, but also such feedback mechanisms "as periodic elections, a vigilant press, petitions to government, and public commentary on proposed administrative regulations" as well as "disruptive demonstrations and legal action."[66] And polycentrists universally recognize the importance of an overarching (typically democratic) governance structure to oversee the experiments that take places at its various centers of decision-making, facilitating information sharing and minimizing negative externalities. The difference between experimental democrats and polycentricists is therefore more a matter of emphasis than

---

[65] Gaus 2016, ch. 4.

[66] Anderson 2010, 99.

anything else. It is best understood as a disagreement over the extent to which progressiveness depends on centralized or decentralized, formal or informal processes—not over whether such processes matters at all.

In any event, I mention these approaches here only to render more concrete what a reasonable level of progressiveness would look like in practice, as well as the sort of questions we must ask about progressiveness more generally. In addition to questions concerning the extent to which progressiveness depends on centralized or decentralized, formal or informal processes, we must also ask about the relative importance of learning and selection mechanisms: do we improve progressiveness primarily by getting better at predicting what will work, or by getting better at stabilizing what has worked and eliminating what has not? Presumably, those who are more optimistic about our predictive capacities will emphasize the former, while those more pessimistic will emphasize the latter—at the limit, abandoning prediction altogether in favor of a pure evolutionary mechanism of random variation and selection.[67] Another salient topic concerns the role of moral diversity

---

[67] The extent to which we can predict the effects of institutional change, and improve these predictions through learning, depends largely on just how complex our world is. An extreme view is that our world is chaotic: the interactions between its features are so dense that minor tweaks to institutions reverberate throughout the entire arrangement in entirely unpredictable ways ("the butterfly effect"). But the relative stability of our world, the history of successful social reform, and the fact that our understanding of institutional functioning has clearly improved across time all suggest that this is not so—that our world, though complex, is "nearly decomposable," such that "the short-run behavior of each [institutional feature] is approximately independent of the short-run behavior of the other components" (Simon 1962, 474; compare Buchanan and Powell 2018, 263-264). Put in these terms, the

in progressive arrangements. Throughout this paper I have simply set aside the issue of what criterion of justice to use when evaluating institutional arrangements by the outcomes they produce. But there is much disagreement about such matters in the real world, and this raises a number of further questions about progressiveness. For example, do progressive institutions require that their members at least form an "overlapping consensus" on a reasonable "political conception of justice"?[68] Or can a diversity of moral and political views—including about the criterion we use for evaluating the outcomes produced by our institutions—coexist with progressive institutions, or even be a driver of progress, as some have recently argued?[69]

Again, I cannot hope to answer such questions here, but only to put them on the agenda. Still, the very fact that they appear to admit a wide range of answers suggests an obvious objection. If it is epistemically infeasible for us to confidently determine what would constitute progress towards the ideally just institutional arrangement, then why isn't it similarly infeasible for us to determine what would constitute progress towards an ideally progressive arrangement? Why isn't my own approach just as epistemically overdemanding as ideal theory?

debate over our predictive capability turns on just how "approximate" this independence is, as well as on how adept we are at identifying the boundaries of approximately independent features. Thankfully, I need not resolve this debate here, since the argument of this paper requires only that features are not so independent that we can confidently identify and track our progress towards the ideal, but not so interdependent that intentional social reform is entirely beyond our ken.

[68] John Rawls 2005.

[69] See, for example, Gaus 2016, Muldoon 2016, and Müller 2019.

Answering this objection provides me with an opportunity to clarify my position. My claim is not that we must identify the ideally progressive institutional arrangement so that we can trade off short-term improvements in justice against progress towards this progressive ideal. It is rather that we must identify short-term improvements in progressiveness, so that we can trade off short-term improvements in justice against short-term improvements in progressiveness. 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 justice, but progressiveness. For example, we might attempt to identify and mitigate biases that feed into our current selection or learning mechanisms—such as the tendency of institutions to change in ways that favor the short-term interests of the rich and powerful rather than long-term justice,[70] or the fact that when we monitor our existing arrangement we often give undue weight to the opinions of some rather than others.[71] Alternatively, we might attempt to solve incentive problems that make our institutional arrangement less conducive to learning and selection—for example, the fact that in most democracies, politicians have an incentive to oversell the benefits of their proposed reforms (to increase their chance of being passed), and then a further incentive to prevent the monitoring of these reforms (because they are unlikely to live up to their bill of goods).[72] Or we might attempt to devise ways to avoid dogmatism and resistance to experimentalism more generally, avoid repeating the mistakes of the past through improved institutional memory and cross-jurisdictional information sharing, and so on.

Of course, there remains the risk that, in only considering short-term justice and

[70] This is a driving concern of Knight and Johnson 2011.

[71] See Fricker 2007 on such "epistemic injustice."

[72] Campbell 1969, 410.

progressiveness in this way, we ultimately make progress away from the ideally just or progressive institutional arrangement. But the severity of this problem is mitigated in two 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 institutional functioning and our ability to make further positive changes, and so generally further rather than set back our ability to improve both justice and progressiveness. And, second, I do not rest my argument on the claim that changing our institutional arrangement in ways we predict will improve progressiveness is in all cases guaranteed to maximally promote long-term justice. No such guarantee is available. Instead, I rest it on the comparative claim that trading off short-term improvements in justice against short-term improvements in progressiveness is a better approach to social reform than either focusing only on short-term justice or attempting to trade off short-term justice against progress towards ideal justice. It represents the appropriate middle ground between the myopia of the former approach and the impracticability of the latter.

5.   Conclusion

Drawing all these threads together, there are, on the approach I have outlined, two basic tasks for theorists of social reform.[73] The first is to engage in problem solving: attempting to identify, to the best of our ability, institutional changes that would ameliorate particular instances of injustice. But, given the complexity of our world and the epistemic limitations it

---

[73] That there are only two basic tasks is compatible with there being other subsidiary tasks that inform these basic ones. For example, one such task may be pure normative theorizing about the appropriate criterion of justice to use when evaluating outcomes. For illustrations of the sort of theorizing I have in mind, see Barrett forthcoming a, forthcoming b.

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, to take a different sort of example, if we live in a democracy whose mechanisms of experimentation, learning, and selection depend on public trust and participation in the democratic process, then we must count it against an unpopular 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 popularity notwithstanding.

The second basic task is to theorize directly about how to improve the progressiveness of our institutional arrangement: the speed and reliability with which it will continue to improve in justice. This, I have argued, depends on its conduciveness to a wide range of promising experiments, to selecting for successful ones, and to learning from both successes and failures. Theorizing about progressiveness differs from problem solving because it is not concerned with ameliorating particular problems of injustice, but rather with improving the progressiveness of the framework within which such problem solving occurs. 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 institutional arrangement more progressive. And we must trade off improvements in short-term justice not against progress towards the ideal, but against progressiveness more generally.

The various questions I have flagged about progressiveness are difficult questions with no easy answers—and we have only scratched the surface of the many issues that

progressiveness raises. But it is precisely these issues to which theorists of social reform must now turn. In a complex world, we cannot assume that ameliorating particular instances of injustice promotes greater long-term justice, but nor can we identify or track our progress towards a long-term goal of ideal justice. So it is only by identifying institutional changes that improve progressiveness that we can figure out how to promote long-term justice. And it is only by implementing such changes that we can effectively pursue it.

Works Cited

Aligica, Paul D. and Vlad Tarko. 2011. "Polycentricity: From Polanyi to Ostrom, and
  Beyond." *Governance* 25(2): 237-262.

Anderson, Elizabeth. 2006. "The Epistemology of Democracy." *Episteme* 3 (1-2): 8-22.

———. *The Imperative of Integration*. 2010. Princeton: Princeton University Press.

Arthur, W. Brian. 2014. *Complexity and the Economy*. New York: Oxford University Press.

Barrett, Jacob. Forthcoming a. "Efficient Inequalities." *Journal of Political Philosophy,*

———. Forthcoming b. "Is Maximin Egalitarian?" *Synthese.*

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 The Courts*, Cambridge: Cambridge University Press.

Buchanan, Allen. 2004. *Justice, Legitimacy, and Self-Determination: Moral Foundations for
  International Law*. New York: Oxford University Press.

Buchanan, Allen and Russell Powell. 2018. *The Evolution of Moral Progress: A Biocultural Theory.*
  New York: Oxford University Press.

Campbell, Donald T. 1969. "Reforms as Experiments," *American Psychologist* 24 (4): 409-429.

Carens, Joseph. 2013. *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.

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. 2009. *Naturalizing Epistemology: Thomas Kuhn and the 'Essential Tension'.* New York: Palgrav MacMillan.

Dewey, John. 1927. *The Public and Its Problems.* New York: Henry Holt and Co.

Estlund, David. 2014. "Utopophobia." *Philosophy & Public Affairs* 42 (2): 207-235.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing.* New York: Oxford University Press.

Gaus, Gerald. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society.* Princeton: Princeton University Press.

Gaus, Gerald. 2018. "The Complexity of a Diverse Moral Order." *The Georgetown Journal of Law and Public Policy* 16: 645-680.

Gilabert, Pablo. 2012. "Comparative Assessments of Justice, Political Feasibility, and Ideal Theory." *Ethical Theory and Moral Practice* 15 (1): 39-56.

Gilabert, Pablo and Holly Lawford-Smith. 2012. "Political Feasibility: A Conceptual Exploration," *Political Studies* 60(4): 809-825.

Hamlin, Alan and Zofia Stemplowska. 2012. "Theory, Ideal Theory, and the Theory of Ideals." *Political Studies Review* 10(1): 48-62.

Hayek, F. A. *The Constitution of Liberty.* 1960. Chicago: University of Chicago Press.

Kitcher, Philip. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions.* New York: Oxford University Press, 1993.

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.

Lawford-Smith, Holly. 2013. "Understanding Political Feasibility," *Journal of Political Philosophy* 21(3): 243-259.

Lipsey, R. G. and Kevin Lancaster. 1956-1957. "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.

Merton, Robert K. 1936. "The Unanticipated Consequences of Purposive Social Action," *American Sociological Review* 1(6): 894-904.

Mill, John Stuart. 1978. *On Liberty.* Indianapolis: Hackett, 1978.

Miller, John H. and Scott E. Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life.* Princeton: Princeton University Press.

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.

———. 2005. *Understanding the Process of Economic Change.* Princeton: Princeton 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. 2018. "The Imperative of Complexity," *Cosmos+Taxis* 5(2): 4-12.

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.

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

———. 2005. *Political Liberalism.* New York: Columbia University Press, 2005.

———. 1999b. *A Theory of Justice.* Cambridge: Belknap Press.

Riker, William R. 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):
341-362.

Rosenberg, Alex. 2016. "On The Very Idea of Ideal Theory in Political Philosophy." *Social
Philosophy and Policy* 33(1-2): 55-75.

Schelling, Thomas. 1971. "Dynamic Models of Segregation," *The Journal of Mathematical
Sociology* 1(2): 143-186.

Schmidtz, David. 2016. "A Realistic Political Ideal." *Social Philosophy and Policy* 33(1-2): 772-
796.

———. 2011. "Nonideal Theory: What it is and What it Needs to Be." *Ethics* 121(4): 772-796.

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

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Design for Generalized Causal Inference.* Boston: Houghton Mifflin.

Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy & Public Affairs* 38(1): 5-36.

Simon, Herbert A. 1962."The Architecture of Complexity." *Proceedings of the American Philosophical Society* 106(6): 467-482.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: Methuen & co, 1776.

Smith, Peter. 1998. *Explaining Chaos.* Cambridge: Cambridge University Press.

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, Phillip E. 2006. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press, 2006.

Valentini, Laura. 2012. "Ideal vs. Non-Ideal Theory: A Conceptual Map," *Philosophy Compass* 7(9): 654-664.

Wiens, David. 2015a. "Against Ideal Guidance." *The Journal of Politics* 77(2): 433-445

———. 2015b. "Political Ideals and the Feasibility Frontier." *Economics and Philosophy* 31(3): 447-477.

———. 2012. "Prescribing Institutions Without Ideal Theory." *The Journal of Political Philosophy* 20(1); 45-70.

Wilson, David S. 2016. "Two Meanings of Complex Adaptive Systems." In D. S Wilson and A. Kirkman (eds.), *Complexity and Evolution: Towards a New Synthesis for Economics.* Cambridge: MIT Press.