

# SOCIAL EXPERIMENTATION IN AN UNJUST WORLD

Jacob Barrett and Allen Buchanan

## Abstract

There is a resurgence of interest in social experimentation as a means of promoting social progress, including progress in justice. In this paper, we first advance an argument in favor of social experimentation drawing on its capacity to resolve uncertainty both about how to achieve socially valuable goals and about which goals are worth pursuing. We then identify four challenges: the information problem (experiments may not yield relevant information), the selection bias problem (potentially informative experiments may not be undertaken), the uptake problem (the information generated by experiments may not be put to good use), and the risk problem (experiments may carry unacceptable risks). Finally, we argue that certain injustices can exacerbate all four problems, rendering social experimentation a less reliable path to progress, and, in cases of severe injustice, perhaps even a regressive force. The upshot is not that we should abandon social experimentation, but that we should temper our expectations and focus on constructing conditions under which experimentation is more likely to be progressive. Specifically, to render social experimentation a more reliable engine for social progress of any sort, we must remedy or mitigate the injustices that diminish its value.

## I. Introduction

Social reform is an uncertain endeavor. The reforms we implement often fail to achieve their desired effects or have unanticipated deleterious effects. And even when attempts at reform work out as planned, we may later revise our standards of evaluation, and so conclude that we did not produce genuine improvements after all.

Reflecting on this uncertainty has led many—from historical figures such as John Stuart Mill

(1977), Karl Popper (2013), and John Dewey (1927), to their legion of contemporary intellectual descendants (e.g., Anderson 2006, Knight and Johnson 2011, Sabel and Zeitlin 2011, Muldoon 2016, Müller 2019, Barrett 2020, Robson 2021)—to endorse an experimentalist orientation toward social improvement.<sup>1</sup> The basic case for experimentalism is elegant and straightforward. Since the institutional or policy alternatives that we currently deem optimal frequently prove otherwise, trying out a variety of promising alternatives often uncovers better options than our current best guess. And even when experiments don't work out, they can provide valuable information about how better to pursue other reforms in the future—either through revealing better ways of implementing our current goals or through helping us to identify better goals. Other things being equal, a society that engages in a wide range of social experimentation should therefore be a more “progressive” society: one that is better at getting better, more conducive to improvement across time (Barrett 2020). This is the thesis of progressive experimentalism.

Our goal in this paper is to cast a critical eye on progressive experimentalism—not in order to refute it, but to take it seriously enough to ask some hard questions. More precisely, our focus will be on how certain injustices can make social experimentation a less reliable path to progress, and, in some cases, may even render social experimentation a regressive force. This is ironic, given that many progressive experimentalists are especially enthusiastic about the ability of social experimentation to promote justice. Our analysis therefore suggests that the relationship between social experimentation and social justice is more complicated than is typically appreciated: in many circumstances where progress in justice is most needed, experimentation is least likely to be a reliable path to improvements of any sort, including improvements in justice.

In Part II, we clarify the ambiguous term “social experimentation.” In Part III we articulate the basic argument that social experimentation promotes progress, including progress in justice. In Part IV, we identify four challenges to this argument: (i) the information problem (experiments may

not yield valuable information), (ii) the selection bias problem (potentially informative experiments may not be undertaken), (iii) the uptake problem (the information generated by experiments may not be put to good use), and (iv) the risk problem (experiments may carry unacceptable risks). Although we will suggest that these challenges serve as a helpful lens for theorizing about social experimentation in general, our primary aim is to show how existing injustices, in particular, can exacerbate all four challenges. This task will occupy us in Part V, where we defend our headline conclusion that certain injustices undermine the progressiveness of social experimentation.

In Part VI, we take stock, drawing out some implications of our analysis. One implication is that, if relevant injustices are too severe, social experimentation may fail to make things better or may even make things worse. But even in reasonably just societies, where experimentation is on balance progressive, injustices can reduce the effectiveness of social experimentation as a means of achieving not only justice, but also other social goods. Although this needn't lead one to abandon a commitment to experimentation, it should encourage more realistic expectations. It also suggests an important practical upshot. To render social experimentation a more reliable engine for social improvements of any sort, it is imperative that we remedy or mitigate the injustices that diminish its value. So, no matter one's normative priorities, there is a strong instrumental case for remedying such injustices in the short term to help achieve those priorities in the long run.

## II. What is Social Experimentation?

Before proceeding, we must clarify terms. We can distinguish between the experimental stance and the deliberate production of experiments. To adopt the experimental stance is to regard our current opinions about the desirability of different policies or institutions as hypotheses that are subject to critical appraisal and possible abandonment or modification in light of empirical evidence—and actively to seek such evidence. One can adopt the experimental stance without engaging in the

deliberate production of experiments, that is, without designing and constructing conditions in which such hypotheses will be confirmed or disconfirmed by evidence—evidence that would not likely be available had those conditions not been constructed.

In this paper, our focus is on the deliberate production of experiments. This is what we mean by “social experimentation.” The thesis of progressive experimentalism is not merely that it is wise to take the experimental stance and view existing and past arrangements as “natural experiments,” but that the deliberate production of experiments fuels social improvement. However, it is worth distinguishing two ways of deliberately producing experiments. First, a society might produce experiments in a direct “top-down” way, as when officials at the highest-level political authority design and impose different experimental policies in different places in order to test and compare their effects. Second, a society might produce experiments in an indirect “bottom-up” way by creating incentives for various parties to engage in experimentation, for example, by delegating authority to different political or administrative bodies in a way that predictably leads them to try out a wide range of policies. Although there are important differences between these approaches, our primary concern is with the deliberate production of experiments in general, rather than any particular regime of social experimentation—whether direct or indirect.

We also emphasize that our focus is not on personal experiments individuals undertake in their own lives but is rather on social experiments—experiments with different policies or institutions. We understand “social experiments” broadly to encompass not only randomized controlled trials but also what are sometimes referred to as “quasi-experiments” because they lack randomization or a control group (e.g., Shadish, Cook, and Campbell 2002). The question we will investigate is whether deliberately producing social experiments is a reliable means of producing social progress. Here, we understand “social progress” broadly to encompass the achievement or enhancement of widely accepted social goods, including justice, as well as well-being, freedom, efficiency, and the like.

### III. The Case for Progressive Experimentalism

The root assumption motivating progressive experimentalism is that achieving progress isn't simply a matter of implementing what we already know to be better states of affairs. Rather, we have much to learn, not only about how to achieve improvements as we currently understand them, but also about what counts as an improvement. For the experimentalist, then, making progress is not simply a matter of putting into practice what we know to be desirable. It is also a matter of finding out what is desirable by engaging in practice—of learning through trial and error.

When we are uncertain, both about what we should seek to achieve and about how to achieve it, we face a tradeoff between exploitation and exploration: between exploiting our current understanding of how to obtain greater value and exploring the space of possibilities to find better ways of doing so. This tradeoff is familiar from daily life. Suppose you are at a restaurant, deciding what to order. Last time, you had the Pad Thai, and you liked it very much. So you are tempted to order it again. But you are unfamiliar with Thai food, and there may be other dishes you would like even more, if only you were to try them. If you always order Pad Thai, you will never discover these options. So you are tempted to try something else instead.

If you settle on the Pad Thai, you are *exploiting* your current information about how likely you are to enjoy various items on the menu. If you try something else, you are *exploring* in the hopes of discovering something better (for example, that you prefer coconut curry) or at least of obtaining information that will help you to make better choices in the future (for example, that peanut sauce is not for you). Although exploiting has greater expected value if we focus only on its immediate effects, in many contexts, people who always exploit get stuck choosing options that, were they to have explored more, they would have forgone in favor of something else. In the long run, they achieve less than they would had they sometimes explored. But people who persist in seeking new information through exploration, while never putting that information to use, also achieve less than they would if

they limited their exploration in order sometimes to exploit what they already know. So exploitation and exploration must be balanced against each other. This is the exploration/exploitation tradeoff.

There is nothing special about this example. The same tradeoff arises in any context where we are uncertain about the value of options and where trying out options can teach us about their value in a way that informs our future choices. A classic illustration is the “multi-armed bandit problem” (Sutton and Barton 2012, ch. 2, provides an overview). You face several slot-machines which you may use some number of times. You want to get as much money as possible but are uncertain about each machine’s cash payoff. What should you do? One approach is always to use the machine you currently believe has the highest expected cash payoff. This amounts to always *exploiting*. But there is a good chance that other machines have an even higher payoff, so you will typically do better sometimes to *explore*: to try out machines you currently expect less from, but which you could discover have a higher payoff upon trying them. Indeed, for a wide range of parameters, agents that play as simple a strategy as exploiting the machine they assign the highest expected cash payoff to most of the time but occasionally choosing some other machine at random tend to get much richer, over the long run, than those who always exploit. More sophisticated strategies—for example, positively correlating the frequency with which one uses a machine with its expected cash payoff—often do even better.

It is important to understand that exploiting doesn’t always mean sticking with the status quo. In the multi-armed bandit problem, even someone using a purely exploitative strategy will switch away from their current machine if, after some disappointing uses, they come to believe that it has a lower expected cash payoff than some alternative. The contrast is not between maintaining the status quo and choosing something new. It is between picking the option your current information suggests is best and trying out other options to generate information about how better to achieve value.

Our basic argument for the long-term benefits of social experimentation, then, is first, that balancing exploration against exploitation produces greater long-run value than always exploiting in

contexts of uncertainty where we can learn through trial and error about the value of the options we will later face; and second, that the pursuit of social improvement is one such context. We now consider an example that illustrates the second point.

Suppose you are a public official with the mandate and authority to tackle some social problem, say, to reduce poverty or ameliorate racial discrimination. You have surveyed the available theories, models, and evidence bearing on various alternatives and now believe that some particular reform would best serve this aim. If your sole goal is to mitigate the social problem, it might seem that what you ought to do is obvious: you should implement that reform wherever possible. But this would be only to exploit, and so may lead you to miss out on better options. If instead you were to explore, to engage in social experimentation—say, by trying out a variety of reforms in a variety of places—you might very well discover that some other reform is better, or at least learn something that leads you to revise your theories or models so that you can come up with better reforms in the future.

Perhaps, for example, your current evidence suggests that a means-tested benefit policy best reduces poverty, but if you give some weight to exploration, you might discover that a basic income policy, say, does much better. Or perhaps your choice depends on your appeal to a model on which providing individuals with unconditional benefits significantly reduces their motivation to work, and social experimentation would allow you to test and calibrate this model, which would in turn facilitate the development of better reforms going forward. More fundamentally, experimentation might lead you to reconceive the problem you are addressing. For example, you might come to believe that the major cause of some form of racial discrimination is institutional and that priority should be given to institutional reforms rather than to educational efforts targeting individuals' prejudices.

In the cases considered so far, what we hope to achieve through exploration is the better realization of antecedently specified goals. In particular, we can obtain either information about how well the specific options we have tried achieve these goals or more general information that bears on

the evaluation of other options we may implement in the future. This alone is enough to demonstrate the long-term value of social experimentation. But it is worth stressing that in social contexts, we similarly face uncertainty regarding the ends we should be striving for, and that exploration can also help us to resolve such uncertainty.<sup>2</sup>

That exploration has this further benefit follows from one further premise, namely, that the full or partial realization of goals in practice provides us with evidence not only about how better to realize them, but also about whether they are indeed worth realizing, how much weight they merit, or what other goals we should pursue instead. This can occur in several ways (compare McGrath 2019, ch. 4). For example, if a social experiment produces an outcome that our current conception of morality registers as an unambiguous improvement, but those living under the experiment credibly protest their new conditions as unjust, their testimony may lead us to rethink our moral stance. Alternatively, a social experiment might produce some new outcome that appears either highly morally desirable or undesirable to us upon reflection, but that we otherwise would not have thought to consider (or whose importance we wouldn't have appreciated when considered in the abstract). It is also common to find that those who live under new arrangements are forced to confront various preconceptions—say, that members of some groups have lower moral status than others, since they are naturally inferior and so would not take advantage of greater educational or economic opportunities even if they were available—and so to revise their moral beliefs accordingly.

We take the occurrence of this sort of moral learning to be uncontroversial. Indeed, we find it plausible that social experimentation can promote learning not only about the mid-level or proximate goals we aim to pursue through policy or institutional design, but also about fundamental moral principles. Notably, even G. A. Cohen, who famously argues that fundamental principles are “fact-insensitive,” agrees that our evaluation of these principles crucially depends on how they play out in actual practice, since “responses to actual facts reveal our principles better than our responses to

hypothesized facts do” (Cohen 2003, 227). Regardless of whether experimentation can teach us about fundamental principles or only mid-level goals, however, the point is that exploration not only helps us better to pursue the goals we accept, but also helps us to understand what goals we should pursue.

There remains an important disanalogy between our examples and real contexts of social reform. In the above cases, choices are made by a single individual, but in actual practice, social change is a complex process over which no individual has complete control. But the exploration/exploitation tradeoff does not only arise in contexts involving a single decision-maker; indeed, there are extensive literatures on how firms balance exploration and exploitation, and, more generally, on how complex adaptive systems strike this balance (beginning with March 1991 and Holland 1992). As these literatures reveal, such contexts provide additional methods for resolving the tradeoff, since they do not require individuals to internalize it. Instead, the tradeoff may be handled at the system level, if some devote their time to exploiting and others to exploring, or if decisionmakers each engage in what they see as exploitation of existing information in a context where this produces exploration because decisionmakers have different information or disagree about what actions their information supports.

Similarly, when it comes to social improvement, we need not think of experimentation as being conducted, say, by a centralized experimenter. Instead, experiments may be overseen by voters, whose representatives engage in experiments in policymaking—say, through widespread use of pilot programs, or the employment of social scientific advisors who help them to design and evaluate policy experiments in ways that better facilitate learning (Campbell 1969). Or, as we noted above, we might achieve social experimentation indirectly, for example, through establishing a “polycentric democracy” in which different political jurisdictions, though each operating under the same overarching democratic framework, serves as its own locus of decision-making about which reforms to experiment with, such that there is widespread variance in which reforms get tried out (Müller 2019). Another option is to establish an “experimentalist governance” framework in which local agencies have

autonomy to implement their own policy experiments in pursuit of common goals, on the condition that they participate in a collective process of subjecting the results of their experiments to peer review and revising them (and their common goals) accordingly (Sabel and Zeitlin 2011).

We raise these three frameworks not to endorse or reject them, but merely to make plausible the claim that social reform is a context, like many others, in which exploration can and should be balanced against exploitation. The key point is that without exploration, we will often fail to discover better options, and will obtain less general information bearing on the evaluation of options we have yet to try. This is the basic argument for progressive experimentalism: social experimentation is an engine for long-term social improvement because exploration generally conduces to long-run value.

#### IV. Four Challenges to Progressive Experimentalism

The basic argument for social experimentation we have sketched is *prima facie* persuasive, but it rests on several assumptions that warrant further scrutiny. In this Part, we outline four such assumptions, explain how each raises a distinct challenge to progressive experimentalism, and illustrate how acknowledging these challenges can enhance further research into social experimentation. In the next Part, we show how certain injustices exacerbate all four challenges.

The first assumption is that we will indeed learn valuable information from social experimentation—information that we can put to use or “exploit” in pursuit of better options going forward. But this is not obvious. Even if I learn something about the relative tastiness of various dishes prepared in one Thai restaurant, this might not generalize to others. And, similarly, social experiments might lack external validity: the fact that a social experiment works well in one context may not provide valuable information about what to do in other contexts. Or perhaps biases at the individual or institutional level will contaminate the interpretation of whatever valuable information experiments would otherwise yield. This is the *information problem*, and it calls into question the value

of social experimentation as an informative form of exploration on which the basic argument for progressive experimentalism relies.

Second, this basic argument presupposes that, in practice, social experiments will be undertaken in a way that genuinely helps us to explore regions of social space relevant to social improvement. But this may not be how things play out. If our dining club's decisions about what to order are determined by majority rule, and if the majority has a bias against food that sounds "ethnic," we may end up only exploring the kids' menu rather than the richer array of options available. And a similar problem may arise in the social case. A society may fail to undertake certain valuable experiments, not because they are less important than others, but for morally arbitrary or pernicious reasons. So even if social experiments can be informative, the sort of experimentation we actually undertake may differ starkly from experimentation that would genuinely inform the pursuit of social progress. This is the problem of *selection bias*—bias regarding which experiments are undertaken.

Third, our argument for social experimentation only succeeds if the information generated through exploration is actually exploited or put to good use in determining what to do. But we cannot assume this. Even if we explore a wide range of options at a restaurant, this may lead us to order food that fails to better satisfy, or even worse satisfies, our preferences in the future, if our decisions about what to eat are determined randomly or by a socially dominant individual whose tastes negatively correlate with the preferences of the rest of the group. And, in the same way, social experimentation may not lead to improvements if the information experiments yield is ignored by those with the power to apply it, and may even lead to regression if the information gathered from it is employed in the pursuit of bad ends. This is the *uptake* problem, and it suggests that even under conditions where social experimentation is genuinely informative, it may fail to be beneficial, or may even make things worse.

Finally, the argument for progressive experimentalism presupposes that the risks of experimentation are not so great that we would be better off avoiding it. But this, too, cannot be

assumed. If I have a life-threatening allergy, and I don't know which dishes might trigger it, then I may be better off sticking with what I know is safe than trying out other options. And likewise, social experimentation might be too risky—for example, because experiments threaten to have negative consequences, to reinforce injustice, or to violate moral requirements. This is the *risk problem*: if the risks are great enough, the benefits of social experimentation may not be worth the costs even when experimentation yields valuable information that is put to good use in promoting social improvement.

Of course, it is one thing to raise these four problems in the abstract, and another to investigate the extent to which they genuinely undermine the progressiveness of experimentation in any given society. To some extent, this will depend on the overarching framework for social experimentation in place. For example, one plausible hypothesis is that more top-down frameworks do better on the information problem, since they may be better able to design and implement experiments in ways that produce valuable and generalizable information, say, by running a range of subtly different experiments in different jurisdictions while leaving others as “controls.” But at the same time, another promising hypothesis is that more “bottom-up” frameworks have an easier time mitigating the selection bias problem: in a polycentric system where diverse jurisdictions each have authority to engage in their own reforms, we will likely have a wider range of experiments, at least if such jurisdictions differ considerably in their decision-making procedures or goals (Müller 2019, ch. 10).

Similarly, when it comes to the uptake problem, competing hypotheses present themselves. On the one hand, more top-down frameworks may have an easier time rapidly implementing policies that experiments have shown to be valuable, remedying an issue found in more bottom-up arrangements where the lack of good information sharing mechanisms and the considerable autonomy of individual jurisdictions may permit them to ignore the results of experiments elsewhere—as often seems to occur, for example, among states in the U.S. federal system (Bell 2015, 134). But on the other, this very capacity may be a curse if the central government is prone to capture by special interest

groups or to pursuing objectionable goals, such as increasingly effective surveillance of its citizens or the personal enrichment of the political class.

Finally, when it comes to the risk problem, the case again seems mixed. Two domain-general risk-reduction heuristics are containability and reversibility (Buchanan 2011, ch. 6). If the effects of an experiment will be limited to one locale, then, other things being equal it is less risky because whatever damage it produces will be relatively limited. If an experiment is reversible—if it can be halted promptly and in such a way that its effects largely cease—then, other things being equal, it is less risky. A plausible hypothesis here is that bottom-up frameworks typically produce more containable experiments (Müller 2019, 183), but that top-down frameworks are better at reversing experiments which are not sufficiently contained and so have damaging negative externalities. If so, it is far from obvious which sort of framework produces less risks on net.

Presumably, the best way to handle these tradeoffs is through sound institutional design. Our own tentative view is that there should be a number of venues with the authority to undertake a wide range of experiments (to mitigate the selection bias problem), but with some overarching institution responsible for reducing the negative externalities of such experiments (to mitigate the risk problem), for gathering information about their outcomes and making it widely available (to mitigate the information problem), and for providing incentives to decisionmakers in various locales to take seriously and use information generated elsewhere (to mitigate the uptake problem). But obviously this is easier said than done. For example, one challenge is to give the central authority sufficient power to perform this coordinating function without empowering it to exert excessive influence on which experiments are undertaken (the selection bias problem), and there is always the risk that the coordinating authority may misrepresent the results of experiments (the information problem), create incentives for their biased uptake (the uptake problem), or ignore or discount risks that experiments may pose to particular jurisdictions or groups (the risk problem). Our point for present purposes is

not to propose any particular institutional arrangement for experimentation, but merely to suggest that meeting the four challenges to social experimentalism may require institutional innovation. This may itself require “meta-experimentation”: experimenting with different frameworks for experimentation.

The value of social experimentation also depends on the availability of alternative information sources about how to improve society. Sometimes, societies may be able to fund more social science research—say, theory and model construction and laboratory experiments—in lieu of engaging in social experiments. Such research typically serves as a complement to social experimentation rather than a substitute, since social experiments are needed to test theories and models and to determine whether the results of laboratory experiments extend to the real world. Yet in some contexts funding more social science research may be superior to experimentalism due to the severity of the information, selection bias, and risk problems: the information value of social experiments might be sufficiently low that the information they provide over and above research isn’t worth the risk.

Perhaps more significantly, societies can sometimes engage in “progressive borrowing” by copying experiments conducted elsewhere instead of engaging in experiments themselves. The analogy here is with the familiar economic phenomenon of “catch up growth”: less-developed countries achieving economic progress through adopting technologies or economic arrangements already in place in other countries, without going through the protracted process by which they were first created.<sup>3</sup> In general, social experimentation tends to be more valuable in societies that are already at the “cutting edge” of successful institutional innovation in some domain than in those that have more “catching up” to do—to again adopt terminology from the literature on economic growth—since the latter societies can rely more on progressive borrowing and thereby avoid the risks of social experimentation. The relative merits of social experimentation and progressive borrowing also depends on the information, selection bias, and risk problems, since these problems drive down the relative benefits and increase the relative costs of social experimentation.

## V. How Injustice Exacerbates all Four Problems

The last few paragraphs represent some first thoughts intended to illustrate some complexities involved in social experimentation and the value of having our four problems in mind when theorizing about it. In the remainder of this paper, however, we narrow our focus to how certain injustices exacerbate all four problems. Although injustices are not the only sources of the problems, we think they are serious and hitherto unappreciated by proponents of experimentalism.

### *The Information Problem*

The primary benefit of social experimentation is epistemic: it can generate information that bears on how effective particular reforms are likely to be, or that is useful for assessing the plausibility of relevant social theories, models, or standards of evaluation (including principles of justice). However, since societies are highly complex, it is typically far from obvious what lessons to draw from an experiment (Gaus 2021, Part III). In general, the effects of an institution or policy depend not on a simple one-to-one causal relation, but on its interaction with other policies or institutions, as well as with other background conditions. So, although an experiment may appear to show how a desired outcome may be produced, that outcome may instead be the result of other factors or may not be achievable through similar means in other contexts (Cartwright and Hardie 2012, Vivaldi 2020).

This difficulty is exacerbated by our proclivity to bias. Observations are theory-dependent, and we are all subject to confirmation bias and motivated reasoning. So, two people with different prior beliefs or interests might disagree about how well an experiment has played out, or even if they agree it has failed or succeeded, they might disagree about why this is so, and so each take it to confirm their prior expectations or self-serving hopes. If they are in the grip of rival ideologies, such disagreements may be even more pronounced—and harder to resolve. The point is that information generated from social experimentation is hard to gather and interpret, rendering misinterpretation

likely given cognitive biases and the distortions that ideologies foster (compare Anderson 2014).

It is here that injustice enters the picture. In the Marxist tradition of ideology critique as developed by the Frankfurt School and by contemporary critical theorists, ideologies function to uphold unjust social orders (Geuss 1981). Although the details of this picture are controversial, the basic idea that less just societies tend to contain more distortive ideologies is commonplace. We need not defend it by reference to the functionalist claim that ideologies flourish *because* they function to support injustice, but might merely note that this correlation appears to hold empirically, perhaps because ideologies paradigmatically represent the status quo as fundamentally just, and it takes more distortion to obscure larger departures from justice. As a result, we should generally expect the problem of ideological bias in the interpretation of the results of experiments to be graver in less just societies. Importantly, however, ideologies do not only include beliefs that disguise or deny existing injustices. They also feature “belief immune systems”: patterns of thinking, and in particular, epistemically flawed cognitive dissonance resolution mechanisms, that cause their adherents to ignore or discount information that contradicts connected beliefs (Buchanan 2020, ch. 8).

If ideologies function to support existing injustices, then one would expect the ideological belief immune system to be especially potent in screening out or discounting information that, if acted on, would challenge existing injustices. This would suggest that in the very circumstances in which experiments are most needed to find ways of reducing injustice, they are likely to be of less value for doing so because the interpretation of their results will be biased toward the perpetuation of injustice. More generally, if ideologies tend to be more prevalent and distortive in less just societies, the risk of experiments being misinterpreted will be greater, other things being equal, in such societies. Such ideological biases can prevent experiments from yielding information that is useful for making improvements of other sorts, not only those regarding justice.

A related problem is that the information value of experiments may be undercut by

“testimonial epistemic injustice”: the testimony of some individuals may be discounted or disregarded entirely, simply because they are identified as members of groups that are regarded as inferior, and this includes their testimony about the effects of social experiments (Fricker 2007). In some cases, members of such groups may be best positioned to make an accurate appraisal of the information yielded by an experiment, yet their voices may be ignored. Or, because they are generally excluded from the predominant practices of knowledge production in their society, they may not have meaningful opportunities to evaluate experiments. In either case, valuable information about the results of an experiment, especially as it impacts disadvantaged groups, may not be made available. Much like ideological distortion, epistemic injustice therefore exacerbates the information problem.

#### *The Selection Bias Problem*

Even if experiments can be undertaken in an informative way, societies that have arrangements in place to encourage experiments and to extract valuable information from them may not undertake the range of experiments needed to yield relevant information about how to remedy injustice or other social problems. This may occur for several reasons, including lack of effective advocates. For example, those most likely to benefit from an experiment and to see the value of undertaking it may, because they are the victims of injustice, lack the political, material, or educational resources needed to conduct the experiment or to mobilize others to conduct it.

Alternatively, an unjust social order may, through the power of the ideology it contains, encourage the victims of injustice either to fail to recognize the injustices they are suffering or to discount their seriousness. This, again, is often seen as the function of ideologies: to uphold an unjust status quo by distorting people’s beliefs and perceptions so that they fail to appreciate injustices. If one fails to recognize an injustice or underestimates its seriousness, then one may not even get to the point of thinking about experiments that might yield knowledge of how to overcome it. Valuable

experiments will not be undertaken because no one sees the need for them. In each case, the problem is that the very injustices that experimentation is supposed to help us overcome limit the value of experimentation by preventing informative experiments from being conducted.

More generally, if the distribution of resources and power in a society is characterized by unjust inequalities, then there will also be injustice in the production of experiments—regardless of whether the experiments themselves pertain to injustice.<sup>4</sup> Those who are systematically disadvantaged by social practices and institutions will also be at a disadvantage when it comes to proposing and gaining support for experiments. The process of social experimentation may be captured by elites who use it to limit experimentation to issues where greater information would serve their goals. An analogous phenomenon is familiar in the case of medical experimentation: resources are disproportionately funneled into research on “diseases of affluence” rather than more widespread and serious medical conditions that are common among the worse off.

Ironically, other things being equal, the selection bias problem is more serious, the greater the need for improvements in the relevant dimensions of justice. This is because the very flaws—and in particular the maldistribution of resources and political power and influence—that are often in most need of reform are likely to prevent valuable experiments from being conducted. However, it is worth pausing to clarify an important point. Although the information and selection bias problems reduce the benefits of social experimentation, the case for progressive experimentalism can succeed even under conditions where these two problems are very imperfectly mitigated. So long as they are engaging in *enough* experiments, and these experiments are informative *enough*, societies that experiment may still be more progressive than those that don't. This assumes, however, that relevant information will be put to good use, and that, even if it is, offsetting costs are not too significant. The uptake problem and the risk problem respectively call these assumptions into question.

*The Uptake Problem*

Recall that the basic case for progressive experimentalism is disarmingly simple: we must balance the exploitation of existing information about how to improve our society against the exploration of alternatives that generate information about how to improve our society going forward. So far, we have considered two worries about the extent to which social experimentation is likely to proceed in an informative way, as the “exploration” side of the tradeoff requires. But there is a further worry about whether such information will be appropriately exploited. Unless the information generated by social experimentation is put to good use—instead of sitting idly or being put to nefarious purposes—social experimentation will not conduce to social improvement.

To appreciate this problem, suppose a society encourages a wide range of social experiments by giving different jurisdictions the authority to try out their own reforms. Suppose also that, through experimentation with different interventions in different jurisdictions, it becomes clear that some genuinely help to improve our society, while others make things worse while promoting the interests of undeserving social elites. This knowledge will only lead to social improvement if it is exploited in pursuit of such improvements. Among other things, this means that there should be some positive correlation between the discovery of better alternatives and their uptake: if experimental reform  $x$  but not  $y$  improves society, then  $x$  should have a greater tendency than  $y$  to spread to other jurisdictions that would also benefit from it. But if the elites in question have inordinate influence on this process, this may result not in reforms that make society better, but instead in interventions that benefit the elites at the expense of social value. In this case, we may even find a negative correlation between the discovery of better alternatives through experimentation and their uptake. Social experimentation may be rendered an engine for sustaining the status quo or regression rather than for progress.

We can put this point in terms of cultural selection. Cultural selection occurs when three conditions are satisfied: there are variants in cultural phenomena, these variants can be transmitted

over time, and there is some mechanism or “filter” that makes some variants more likely to transmit than others. With some simplification, we can say that cultural selection occurs when some variants outcompete others because they are more amenable to transmission according to existing mechanisms of transmission (Richerson and Boyd 2005 provide an overview). Although social experimentation might seem to speed up social improvement by introducing more variants on which cultural selection can operate, the worry is that unjust power inequalities may introduce a severe bias into our mechanisms of cultural transmission from the standpoint of what we ought to value. Policies may be sustained and spread because they are favored by those who have disproportionate control over public policy, educational practices, or the deployment of information technologies.

Certain injustices may therefore undermine our tendency to exploit information about how well particular policies work in ways that are conducive to social improvement. A related problem concerns the exploitation of general information gleaned from social experimentation about the likely effects of other policies. The issue is that such information is typically as useful to those who wish to improve their position at the expense of others as to those who wish to use it to improve society. The rich and powerful may capture the policy process and use this information to line their pockets or to further entrench their power. Because such information is itself a valuable resource, social elites may also have greater access to it, and may be able to successfully disguise such policies as attempts to promote socially valuable ends.

The uptake problem is not uniquely a problem for progressive experimentalism. It is a more general problem about how information is put to use in society, regardless of its source. However, we emphasize this problem for two reasons. First, as we have noted, in cases where the uptake problem is sufficiently severe, it not only reduces the value of experimentation, but can even render experimentation detrimental. The problem is not merely that information generated through experimentation may fail to be put to good use, but that it may be put to bad use. Second, it is all too

tempting to think that in the face of a problem we are unsure how to solve, we simply need to try out various alternatives to find out which works best. The uptake problem reminds us that a crucial step is missing in this picture: someone must decide which alternative to implement, and they will not necessarily pick the best option. There are many familiar worries about how, when power is unjustly distributed, those with greater power can disproportionately influence the political process. None of these worries go away simply once we adopt an experimentalist approach.

### *The Risk Problem*

In societies that adequately mitigate the information problem, the selection bias problem, and the uptake problem, social experimentation has important benefits: it generates information about how to improve society that is exploited to this end. But this does not yet show that social experimentation is on balance progressive, because it may generate costs that outweigh these benefits. Here, we do not merely have in mind the tradeoff at the heart of the case for progressive experimentalism: that exploration involves the short-term opportunity cost of forgone exploitation. Rather, we are concerned that experimentation may bring risks that cumulatively outweigh the potential benefits of such experiments, even in the long run.

The risk problem is challenging in part because solving it requires more than resolving to refrain from excessively risky experiments. The risks of social experimentation—both the magnitude of possible harms and the probability that any such harm will occur—are typically difficult to ascertain, even by the most knowledgeable and well-motivated experts, due to the complexity of the phenomena involved and the limits of our knowledge. Even genuine experts with the best of intentions may err in estimating risks. An unacceptably risky experiment might therefore be undertaken because information about the risk of harmful consequences was not available or was not heeded; hence the problem of unintended bad consequences. But there are other sources of faulty risk assessment. For

example, the risks of an experiment might be underestimated or overlooked by those who propose an experiment and expend political capital mobilizing support for it, because confirmation bias or motivated reasoning leads them to focus too much on favorable rather than unfavorable outcomes.

This problem is exacerbated by injustice because, in many cases, the benefits and risks of an experiment are unevenly distributed (Knight and Johnson 2011, 48-49). If those who have the most influence over whether an experiment is undertaken have the most to gain if it works out, or the least to lose if things go wrong, the result may be an underestimation or discounting of risk. In societies characterized by unjust power inequalities and epistemic injustice, the predictable result will be communities lacking political or epistemic influence disproportionately bearing the costs of social experiments. So existing injustices may result in the undertaking of overly risky experiments, or in experimentation that perpetuates injustice through unjustly distributed risks.

Another aspect of the risk problem is that social experiments might carry moral risks in the sense that they risk violating independent moral requirements. In the case of medical experimentation, there is widespread agreement that informed consent is necessary to justify an experiment—but such consent may be hard to come by for social experiments. A convincing defense of social experimentation would either have to accept the informed consent requirement and acknowledge corresponding limitations on social experimentation; to show that such limitations are not as severe as they might seem, perhaps because certain regimes of social experimentation allow individuals to opt-out of experiments by moving away from jurisdictions where they are occurring (Müller 2019, 176-177); or to show why informed consent is required for medical but not social experiments (compare MacKay and Chakrabarti 2019). One promising line of argument for this last conclusion is that democratic processes can take the place of consent, when it comes to institutional or policy decisions. Indeed, democracy can be seen as a mechanism for achieving collective decisions under conditions where the requirement of unanimous consent would be unworkable.

Regardless of how one handles this issue, the key point is that moral risks will generally be greater, as well as less justly distributed, in the presence of the injustices we have been focusing on. This will be true given any of the three defenses of social experimentation we have mentioned: in the presence of unjust power inequalities, epistemic injustice, and ideological distortion, there is a greater risk that those with less political or epistemic resources will be experimented on without their informed consent; such individuals will be less likely to have the meaningful ability to opt-out of experiments (for example, since they are less able to move); and the democratic process will be less likely to adequately take their interest or opinions into account in a way that would legitimize experimentation. This is yet another way that, in the presence of certain injustices, social experimentation risks perpetuating rather than mitigating injustice.

## VI. Implications and Conclusion

Recently, political philosophers have exhibited heightened interest in social experimentation as an engine for progress. While we have argued that social experimentation can serve as a form of exploration that generates valuable information about how to improve society, we have also argued that it is subject to four major challenges: the information problem, the selection bias problem, the uptake problem, and the risk problem. The first two problems call into question how much valuable information experimentation generates in practice, the third whether this information is exploited or put to good use in pursuit of social improvement, and the fourth whether the benefits of exploiting such information outweigh the costs of experimentation. We noted that the four challenges provide a helpful framing for future research into social experimentation, but then narrowed our focus to how certain injustices—specifically, epistemic injustice, unjust inequalities in power and influence, and the injustices associated with ideological distortion—may exacerbate all four problems. If we assume (as seems plausible) that injustices tend to cluster together, then the implication is that, in less just

societies, we should generally expect social experimentation to generate less valuable information, for the information it generates to be less likely to be put to good use, and for experiments to impose greater and less justly distributed countervailing risks.

While we have identified several plausible mechanisms by which injustices can exacerbate the four problems, the extent to which these mechanisms are genuinely operative in existing societies is a largely empirical question that requires further study. We will not attempt to answer this question here. Instead, we end by suggesting some general lessons to draw from our discussion about the role of social experimentation in promoting social progress.

One upshot of our argument is that some societies may be characterized by such deep injustices that social experimentation is detrimental in them. This can happen in two situations, depending on the severity of the uptake problem. In the first, where the uptake problem is less severe, the information generated through social experimentation still tends, on balance, to be put to good use. But injustices increase the risks of experimentation (the risk problem) and reduce its benefits (by exacerbating the other three problems), to the point that the risks outweigh the benefits. This is a case where social experimentation has some benefits, but the benefits are not worth the costs. Second, regardless of the other problems, in the presence of severe injustice, the uptake problem may, all by itself, render social experimentation detrimental: information gathered through experimentation may tend, on balance, to be put to bad use. The political process may be so driven by individuals pursuing goals that negatively correlate with justice and other social goods that additional information generated by social experimentation would only serve to help these individuals further entrench their power or better pursue these deleterious goals.

In either situation, the lesson is clear. If social experimentation is to be a path toward progress, the path must first be cleared by other means. But it makes a significant difference which of the two situations we are in. In a society where social experimentation is detrimental due to its serious risks,

but the uptake of information still, on balance, tends toward social improvement, it may be possible to achieve progress through adopting a purely exploitative (as opposed to exploratory) approach to reform. This is especially plausible if the society can engage in what we earlier called “progressive borrowing”: if politically powerful members of an unjust society have sufficient incentives to appropriate the results of experiments undertaken elsewhere, they might reap the benefits of experimentation without conducting their own experiments. This might occur, for example, if leaders of these societies have incentives to improve their society to enhance their prestige or in response to external pressure, or if such leaders genuinely wish to improve things. It is only once enough progress has been made that the benefits of social experimentation outweigh its risks that social experimentation should begin. Thus, a key question is whether there are currently enough examples of societies who have achieved this level of progress, and in ways that are generalizable and so can be fruitfully copied, that societies in less favorable conditions can similarly achieve this through progressive borrowing. Reforms that bring society to this point are especially valuable in the long run, since they allow societies to tap into the progressive engine of social experimentation.

The second situation is more troubling. If the uptake problem is so severe that the information social experimentation generates is on balance likely to be put to bad use, then progressive borrowing cannot come to the rescue: information gathered from experiments elsewhere will likely be put to bad use as well. This may most obviously happen in a dictatorship, but it might also happen in a society pursuing a pernicious conception of morality or justice or whose political process has been captured by private interests. In this situation, it may take an exogenous shock, or a strike of good luck, to get things on track. Alternatively, the best hope may be change initiated by an enlightened reform group that has the power to effect change and has somehow risen above the biases that are pervasive in their society. This will often be a disadvantaged group that suffers a severe injustice and so best understands and is most motivated to rectify it, and who is able to persuade some members of the dominant group

to adopt or sympathize with their cause. The American civil rights movement following the U.S. Supreme Court ruling in *Brown v. Board of Education* might be an example of this kind. Where such changes significantly reduce the injustices that undercut the value of experimentation it may then be appropriate to rely on experimentation as a means of achieving further progress.

Thankfully, many societies are not so thoroughly unjust. And in reasonably just societies, where the injustices at issue are not as deep or pervasive, social experimentation may remain on balance progressive, despite the various problems we have raised. However, even under these more favorable circumstances, injustices may significantly reduce the value of social experimentation. Here, the general upshot of our argument is not that we should give up on experimentation altogether, but that we should moderate our expectations about how well social experimentation promotes progress. As we have emphasized, there is always a tradeoff between exploration and exploitation, and in less just societies, it may often make sense to tilt the balance further toward exploitation.

Social experimentation may also be progressive in some domains though not others within the same society. For instance, in some societies experimentation with different schemes of anti-discrimination legislation may be beneficial, though experimentation with different economic policies is not—say, because powerful corporations influence the uptake of economic experiments in a way that served their interests but have little reason to negatively influence the uptake of experiments with anti-discrimination policies. Or the opposite may be true in a society that is deeply racist yet committed to economic efficiency or economic justice. Similarly, it may be that experiments with informal institutions or social norms can proceed in a progressive way in some societies whose formal political process has been so captured that experiments with formal institutions or policies would be unwise.

An additional implication of our analysis, then, is that, when proposing social experiments or designing or reforming experimental frameworks, we should think carefully about how social experimentation can go wrong in the presence of injustice, and whether it might be possible to mitigate

the problems we have raised. For instance, in a context where epistemic injustice is especially potent, we might require that those charged with evaluating experiments are drawn from diverse groups of society.<sup>5</sup> Or we might employ targeted techniques, such as the “snowball approach,” whereby those tasked with proposing or evaluating an experiment ask not only all stakeholders known to them for input, but also ask those stakeholders who else they think should be consulted, and then ask those recommended who else they think should be consulted, and so on, until, hopefully, all relevant stakeholders are represented (Rossi, Lipsey, and Freeman 2003, 87, 124).

Similarly, we might try to limit the impact of inequalities in wealth and political influence on biasing the selection of experiments. For example, this might be ameliorated by special subsidies, provided by philanthropic organizations, to be awarded to groups that propose experiments that would otherwise not be adequately funded (compare Reich 2018). Priority might be given to groups whose lack of resources is plausibly due to injustice, since they may provide new information about how progress is achieved that would likely be unavailable without their input.

In the end, however, such mitigating strategies only go so far. For instance, significantly reducing the negative impact of epistemic injustice on the information value of experimentation ultimately requires us to ameliorate the larger systemic injustices that foster epistemic injustice, and significantly reducing the negative effect of unjust power inequalities on the selection of experiments or the uptake of information requires us to reduce those inequalities. Doing so may require major changes in the structure of society.

This brings us to a final implication. If one accepts the progressive experimentalist argument that social experimentation can significantly promote social improvement not only in justice but also in other important domains, and if one accepts our argument that injustice can reduce the rate or reliability at which social experimentation has this beneficial effect, then one should recognize the long-run instrumental importance of ameliorating social injustice in the short term. Specifically, our

argument suggests that ameliorating present injustices can make social experimentation a more reliable or faster engine for promoting socially valuable outcomes, and so can help us to achieve whatever social values we have—whether they pertain to justice or not—in the long run.

If this is right, then our paper can be seen as providing an argument for the following convergence between the amelioration of certain injustices and the pursuit of long-run value: one important way to promote social value in the long term is to reduce these injustices in the short term. To the extent that social experimentation has the potential to promote long-term social progress, we should be especially concerned with ameliorating injustice in order better to achieve this potential.<sup>6</sup>

#### Works Cited

- Anderson, Elizabeth. 2006. The Epistemology of Democracy. *Episteme* 3: 8-22.
- Anderson, Elizabeth. 2014. The Quest for Free Labor: Pragmatism and Experiments in Emancipation. *The Amberst Lecture in Philosophy* 9: 1-44.
- Banerjee, Abhijit and Esther Duflo. 2012. *Poor Economics*. New York: Public Affairs.
- Barrett, Jacob. 2020. Social Reform in a Complex World. *Journal of Ethics and Social Philosophy* 17: 103-132.
- Bell, Daniel A. 2015. *The China Model*. Princeton: Princeton University Press.
- Buchanan, Allen. 2011. *Beyond Humanity?* Oxford: Oxford University Press.
- Buchanan, Allen. 2020. *Our Moral Fate*. Cambridge: MIT Press.
- Campbell, Donald T. 1969. Reforms as Experiments. *American Psychologist* 24: 409–429.
- Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy*. Oxford: Oxford University Press.
- Dewey, John. 1927. *The Public and its Problems*. New York: Henry Holt and Company.
- Fricke, Miranda. 2007. *Epistemic Injustice*. Oxford: Oxford University Press.
- Gaus, Gerald. 2021. *The Open Society and its Complexities*. Oxford: Oxford University Press.

- Geuss, Raymond. 1981. *The Idea of a Critical Theory*. Cambridge: Cambridge University Press.
- Holland, John H. 1992. *Adaptation in Natural and Artificial Systems*. Cambridge: MIT Press.
- Knight, Jack and James Johnson. 2011. *The Priority of Democracy*. Princeton: Princeton University Press.
- Mackay, Douglas and Averi Chakrabarti. 2019. Government Policy Experiments and Informed Consent. *Public Health Ethics* 12: 188-201.
- March, James G. 1991. Exploration and Exploitation in Organizational Learning. *Organizational Science* 2: 71-87.
- Mill, John Stuart. 1977. *On Liberty*, in J. M. Robson (ed.) *The Collected Works of John Stuart Mill*, vol. 18. Toronto: University of Toronto Press.
- Muldoon, Ryan. 2016. *Social Contract Theory for a Diverse World*. New York: Routledge.
- Müller, Julian F. 2019. *Political Pluralism, Disagreement, and Justice*. New York: Routledge.
- Page, Scott E. 2007. *The Difference*. Princeton: Princeton University Press.
- Popper, Karl. 2013. *The Open Society and its Enemies*. Princeton: Princeton University Press.
- Reich, Robert. 2018. *Just Giving*. Princeton: Princeton University Press.
- Richerson, Peter J. and Robert Boyd. 2005. *Not by Genes Alone*. Chicago: University of Chicago Press.
- Robson, Gregory. 2021. The Rationality of Political Experimentation. *Politics, Philosophy & Economics* 20: 67-98.
- Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. 2003. *Evaluation*, 7<sup>th</sup> ed. London: Sage Publications.
- Sabel, Charles F. and Jonathan Zeitlin. 2012. Experimentalist Governance. In D. Levi Faur (ed.) *The Oxford Handbook of Governance*. Oxford: Oxford University Press.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Sutton, Richard S. and Andrew G. Barto. 2018. *Reinforcement Learning*, 2<sup>nd</sup> ed. Cambridge: MIT Press.

Vivalt, Eva. 2020. How Much Can We Generalize from Impact Evaluations? *Journal of the European Economic Association* 18: 3045-3089.

---

<sup>1</sup> Though not our focus here, related approaches have also become popular in economics. For example, see Banerjee and Duflo 2012, who (with Michael Kremer) won the 2019 Nobel Prize in Economics “for their experimental approach to alleviating global poverty.”

<sup>2</sup> This may be a point at which Popper parts ways with Mill and Dewey. For Popper, the primary benefit of social experimentation is the generation of instrumental knowledge; for Mill and Dewey, experimentation also helps us to revise our ends. Thanks to Joshua Forstenzer for helpful discussion.

<sup>3</sup> Thanks to an anonymous referee for suggesting this analogy.

<sup>4</sup> As we will argue, power inequalities also exacerbate the uptake problem. This echoes a common objection to Dewey: that his experimental approach to democracy fails to take power into account. Whether Dewey himself was so blind to power is a subject of ongoing debate. For an explicitly Deweyan approach that aims to accommodate this concern, see, Knight and Johnson 2011, who argue that departures from “equal opportunity of political influence” undermine the experimental process.

<sup>5</sup> Every contemporary proponent of experimentalism we cite references the importance of including diverse individuals in the experimental process. On the epistemic benefits of diversity, see Page 2007.

<sup>6</sup> For helpful comments and discussion, we thank Samuel Bagg, Fred D’Agostino, Jamie Draper, Hilary Greaves, Mario Juarez-Garcia, Andreas Mogensen, Ryan Muldoon, Julian Müller, Sarah Raskoff, Gregory Robson, Simone Sepe, Orri Stefánsson, Christian Tarsney, Teruji Thomas, an anonymous referee, and participants in the Moral Science Research Group at the University of Arizona, in the Global Priorities Institute Works in Progress Group at the University of Oxford, and in the 9<sup>th</sup> Annual Workshop for Oxford Studies in Political Philosophy.