A Little Reality is a Dangerous Thing

Justin Weinberg
University of South Carolina
jweinberg@sc.edu

DRAFT

1. Introduction

How realistic should normative political philosophy be? Our choices fall along a spectrum. At one end is utopianism. We could set out a theory for the truly just society inhabited by ideal persons under highly favorable circumstances. Yet if we are not ideal persons and do not inhabit highly favorable circumstances, it is unlikely that such a theory would be feasible, or, if it were feasible, that it would give *us* particularly good advice. At the other end of the spectrum is a kind of strict realism. We could set out a theory that sticks close to recommending a society we are sure is possible, namely, our own. Yet few of us believe the distance between where we are and where we should be is so small. A theory that did little more than endorse the status quo would be normatively overmodest.

The consensus on this question for many years was thought to reside somewhere between these extremes, with John Rawls's idea of a "realistic utopia."[1] A realistically utopian theory holds out the prospect of improvement for us and our social world, while constraining that prospect to what we think is possible given what we know about us and our world. There are obviously some questions about what this means, but suffice it to say that Rawls took himself to be offering a realistically utopian theory.

Part of what makes a theory realistically *utopian* is that it is an exercise in what we have come to call, following Rawls, *ideal theory*.[2] To engage in ideal theory is to make idealizing assumptions about the behavior and attitudes of individuals and the design and functioning of institutions. There are varieties of ideal theory, but typically it proceeds by determining which principles of justice would be appropriate for a society in which it is more or less assumed that (a) individuals comply with the requirements of justice in their everyday lives, (b) individuals actively support with their attitudes and actions the principles of justice and the institutions the principles require, (c) institutions which satisfy the principles of justice will be in place, and (d) these institutions function well not by luck but by design, and are not marred by other serious deficiencies.

Recently ideal theory has been subject to a number of criticisms which suggest it is closer to the utopian end of the spectrum than previously thought. Chief among these has been what I call the *bad effects* criticism: were the principles of justice identified under ideal theory advocated or implemented in our non-ideal world, the results would be suboptimal, if not downright disastrous. Those who advance the bad effects criticism argue for a turn towards *non-ideal theory*, which attempts to address questions in political philosophy without resorting to idealizing assumptions. It suggests replacing these

---

[1] A theory of justice is realistically utopian "when it extends what are ordinarily thought to be the limits of practicable political possibility and, in doing so, reconciles us to our political and social condition." John Rawls, *The Law of Peoples*, (Cambridge, Mass.: Harvard University Press, 1999), p. 11.

[2] John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, Mass.: Harvard University Press, 1999), pp 7-8. See also p.4 on well-ordered society, p. 125 on strict compliance; and pp. 215ff on non-ideal theory.

assumptions with empirical social science. Its aim is to avoid the bad effects criticism by being more realistic.

This will not work, or so I argue in this paper. Non-ideal theory emphasizes reality, but if we are *even more realistic* than it calls for, we can see that non-ideal theory is also subject to the bad effects criticism. Mainly this is because it risks preserving features of our lives and world whose badness we tend to carelessly overlook and whose existence we unwarrantedly assume as static and permanent, a claim I support with help from social psychology, behavioral economics, and history.

That the bad effects criticism applies to non-ideal theory tells us something about ideal theory, or so I will explain. It provides both reinforcement for a traditional role for ideal theory as a kind of external review, and direction in conceiving of a new role for it, too, as a program for experimentation. As I'll explain, the success of ideal theory in these roles does not depend on avoiding bad effects.

The structure of the paper is as follows. In Sections 2 and 3, I introduce the bad effects criticism of ideal theory and the move toward non-ideal theory. Non-ideal theory seeks to replace idealizations with an empirically accurate *description* of us and the world, and an empirically accurate *prediction* about what is possible for beings like us in a world like ours. In Section 4, I provide reasons to doubt our capacity to provide sufficiently accurate descriptions of how things are. In Section 5, I provide reasons to doubt our capacity for accurately predicting how things could be. Together, these doubts form the basis of an argument that shows that non-ideal theory is subject to the same criticism—bad effects—as ideal theory. Sections 6 and 7 propose roles for ideal theory that learn from non-ideal theory's mistakes.

2. Ideal Theory and Bad Effects

The bad effects criticism of ideal theory rests on a presumption that theories of justice are supposed to be action guiding: theories of justice give us principles that we are to accept and use as the basis for political, economic, and social reorganization. This is not an outlandish presumption. Such theories purport to be telling us what justice requires, and if justice requires something of us, that seems to be a very strong reason to do that thing.[3] (Some philosophers, such as David Estlund, defend ideal theory in part by denying that this is necessarily the case.[4] I briefly return to the question of the action-guiding aspect of political philosophy later in the paper.)

If political theories are supposed to be realizable, they can be critiqued when they are not, and that is precisely the criticism that has been launched against ideal theory. The criticism is that by assuming away particular details or social problems, ideal theory results in prescriptions that are problematic for our non-ideal world.[5]

---

[3] But it is not simply the normativity of principles of justice that lend support to the idea that such principles are to be *realized*. Realization seems to be a main concern of political theorists. The most famous example of this is Rawls's reworking of his theory around the idea of "the fact of reasonable pluralism". See, for example, John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), p.p.36-37.

[4] See David Estlund, *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press, 2008), ch. 10. Also see G. A. Cohen, *Rescuing Justice & Equality* (Cambridge, Mass.: Harvard University Press, 2008).

[5] Now as a universal charge this criticism must fail: not all idealizations are obstacles for all prescriptions. (*Homo economicus* is an idealization and there are plenty of cases in which its simplification of human nature does not render policies based on it unsuccessful in the real world.)

Liam Murphy puts the point this way: "An acceptable theory of justice must have acceptable implications for both ideal and nonideal theory."[6] He then goes on to point out a feature of Rawls's ideal theory that he claims fails this test. Murphy critiques Rawls's view of what the principles of justice apply to. According to Rawls, the principles of justice apply to the institutions of the basic structure of society, not to private individuals in their everyday lives.[7] This is usually interpreted to mean that the justness of a society is a function of its basic institutional structure. Murphy's critique of this idea is based on the point that sometimes private individuals will be in a better position to directly promote the ends of the principles of justice themselves, such as by giving to a humanitarian aid agency, rather than to act in ways that are mediated by the basic structure—such as reforming or creating new institutions.[8] Rawls's view, Murphy claims, cannot recognize the justice-serving capacities of ordinary individuals.

In our world, just institutions are absent and institutional reform is hard work. Justice would be better served by a theory that was compatible with private individuals themselves at least some of the time being directly responsible for the production of justice. The idealizing assumptions built into Rawls's conception of a well-ordered society—particularly regarding the presence or ready availability of just institutions—lead him to construct a theory of justice that risks leaving the world worse off, in terms of justice, than it otherwise would be. "If our theory has implausible implications for the nonideal case, the theory may have some intellectual interest, but it would fail as a normative political theory."[9]

Putting aside whether Murphy's critique is sound, the form is clear: ideal theory embeds certain assumptions that render it likely to bring about bad effects were it implemented.

Colin Farrelly takes a different route to a similar conclusion. In his critique of ideal theory, Farrelly argues that ideal theorists fail to take into account the costs of rights, and thus overlook both a source of the scarcity of resources and the effect this scarcity has on

---

[6] Liam Murphy, "Institutions and the Demands of Justice," *Philosophy and Public Affairs* 27, no.4 (Fall 1998), p. 278. This claim, call it *acceptable implications*, sounds quite reasonable, though it needs specification. Murphy is saying that an acceptable theory of justice must have acceptable implications for nonideal circumstances. Yet this is too strong a requirement, for it does not specify the kind of nonideal circumstances we are to consider. A theory of justice may indeed give a proper account of what should be done in *some* nonideal circumstances. But the theory could still be rejected on the basis of *acceptable implications* because it is possible to come up with new, more challenging nonideal circumstances for which the theory has unacceptable implications. This seems true of any theory. For each theory we could come up with progressively worse and worse (unjust, nonideal) scenarios that take us further and further from the ideal circumstances for which the theory is primarily intended. In that case, *no* theory would meet the requirement. And if no theory would meet the requirement, then no theory of justice would be acceptable. But certainly *some* theory of justice is acceptable. Therefore, the requirement is too strong.

This objection could probably be avoided by specifying the range of nonideal circumstances to which a theory of justice must be adequately responsive. Though I will not attempt this work here, it does not seem impossible. Even if we think that only a fuzzy line could be drawn between the relevant and irrelevant nonideal circumstances, we can say that there are some kinds of nonideal circumstances to which it is reasonable to expect theories of justice to apply. After all, the fuzzy line between day and night does not keep us from turning on lights when we need them to find our way.

[7] See Murphy, "Institutions and the Demands of Justice," and, for example, Rawls, *Political Liberalism*, pp.268-69.

[8] Murphy 1999, 281. He provides an example of global injustice and writes about Rawls that he "would believe… that justice requires an egalitarian set of institutions to replace the mostly informal and decidedly inegalitarian institutions that currently prevail. But it could not be right that an individual rich First Worlder is required to devote her resources to the Quixotic task of promoting just international institutions. Such a person could clearly do so much more to alleviate suffering or inequality by doing what she can on her own—by giving money to humanitarian aid agencies."

[9] Murphy, "Institutions and the Demands of Justice," p. 279.

the reasonability of tradeoffs between the protection of different rights. One of his targets is Rawls's theory, particularly Rawls's view that the liberties specified by his first principle of justice are not to be sacrificed for the sake of meeting the distributional requirements specified by the second principle.

Rawls's "cost-blind"[10] approach might work just fine, Farrelly says, if we lived in a society sufficiently developed to maintain a democracy *and* if the liberties specified by the first principle of justice never needed defense or maintenance—that is, if we lived in fairly ideal circumstances. However, when either or both of these conditions fail to obtain, the liberties of the first principle become very difficult or costly to protect. We might find ourselves in situations in which a majority, if not the entirety, of a society's economic resources are directed by Rawls's theory into trying for minuscule improvements in securing people the liberties specified by the first principle, when such resources could have done everyone much more good had they instead been put to the kinds of aims associated with the second principle (such as reducing unfair economic hardship).

Farrelly writes, "if Rawls' theory is supposed to yield principles of justice that can serve as a guide for the collective action of citizens in open, partially compliant societies, then Rawls' 'simplifying' assumptions will prove problematic."[11] This echoes Murphy's complaint. Both theorists are concerned about the problems that might arise were theories constructed atop idealizing assumptions implemented in non-ideal circumstances, i.e., the bad effects criticism.

3. The Turn to Non-Ideal Theory

If the bad effects criticism of a political theory is warranted because of idealizing assumptions that take the place of real world information about our non-ideal world, a natural response is to incorporate such information into political theory.

So instead of those assumptions, we would have information about (a) average rates of compliance with various laws and policies, (b) people's attitudes about justice and their participation in opportunities to act in support of the principles of justice, (c) which institutions it is possible for us to bring about, and (d) what empirical contingencies such institutions depend upon for their creation and functioning, and what their side effects are. Farrelly urges theorists to take into account "the common-sense facts of political sociology."[12] Presumably similar categories of common-sense facts relating to economics, law, psychology, biology, and so on, are also relevant.

In short, the non-ideal theorist says we should use what we know about us and our world so as to craft theories of justice that are "realistic".[13,14]

---

[10] Colin Farrelly, "Justice in Ideal Theory: A Refutation," *Political Theory* 55, no.4, p. 845.
[11] Farrelly, "Justice in Ideal Theory," p. 850.
[12] Farrelly, "Justice in Ideal Theory," p. 852.
[13] Farrelly, "Justice in Ideal Theory," p. 853.
[14] This criterion of realism has to be cashed out in a way that does not put it at odds with the normativity of political theory, which is a point made in numerous works. Recent examples include: Mark Jensen, "The Limits of Practical Possibility," *Journal of Political Philosophy* 17, no.2 (June 2009), pp. 168-84; Laura Valentini, "On the Apparent Paradox of Ideal Theory," *Journal of Political Philosophy* 17, no.2 (September 2009), pp. 332-55; Zofia Stemplowska, "What's Ideal About Ideal Theory?" *Social Theory and Practice* 34, no.3 (July 2008) pp. 319-40. We know that the status quo is achievable, but since the status quo is not necessarily just, a theory of justice could not be limited to endorsing the status quo. We need a concept of the practically possible to work with (Jensen). And then we need an account of the extension

4. Non-ideal theory and the Status Quo

If political philosophy can be interpreted as bridging the gap between justice and the real world, the difference between ideal and non-ideal theory is which side of the bridge we are starting at. Ideal theory starts on the justice side, and non-ideal theory starts in the real world.

Starting in the real world means starting with problems for our theory to solve and obstacles for our theory to acknowledge. For example, Norman Daniels, a sympathetic critic of Rawls, reminds us with his work that in the real world, people get sick. Susan Okin reminds us that people have families. Farrelly reminds us that rights have costs. Murphy draws our attention to the possibility of institutional failure. Charles Mills argues that ideal theory overlooks the significance of race. Ingrid Robeyns directs us to take into account the possibility of unintended consequences of the actions based on our principles and policies. Ilya Somin argues that the problems of voter ignorance and state autonomy (the capacity of the state to ignore the will of the people) are problems that should not be, but are, largely ignored by political theorists. And so on.[15]

Non-ideal theory takes the world as it is and people as they are as building materials in its theory construction. Quality construction depends on quality materials: it is important to the success of non-ideal theory that the information it makes use of is accurate. Yet, as I will explain in this section of the paper, we have good reasons for doubt about our accuracy.

These reasons include:

a. Status Quo Bias. It is a well-confirmed result of studies in social psychology and behavioral economics that people exhibit a cognitive bias in favor of the status quo. In comparing the desirability of outcomes, people have a tendency to prefer the status quo simply because it is the status quo.[16]

If people suffer from status quo bias, they will tend to give a more favorable accounting of their current circumstances than might be warranted by an impartial perspective. Elements of society that we are familiar with may seem less objectionable to us than they would to outsiders (or to us, if we lacked the bias), since these elements are part of our status quo. This poses a difficulty for any normative theorizing, but it is especially problematic for non-ideal theory, since non-ideal theory, with its emphasis on the world as it actually is, draws our attention to various elements of the status quo in a way that more ideal theorizing does not. The ideal theorist starts on the other side of the bridge. She may embark upon her project by picturing what an ideally just state or society would be like, not by picturing the status quo. Non-ideal theory, in contrast, begins with a focus on issues and aspects of the world that are likely to be strongly under the influence of

---

of that concept. It is with this filling in of the details that we will see that non-ideal theory runs into a version of the bad-effects criticism.

[15] See Daniels, *Just Health Care* (Cambridge: Cambridge University Press, 1985); Okin, *Justice, Gender, and the Family* (New York: Basic Books, 1991); Farrelly, "Justice in Ideal Theory"; Murphy, "Institutions and the Demands of Justice"; Mills, "'Ideal Theory' as Ideology," *Hypatia* 20, no.3 (2005); Robeyns, "Ideal Theory in Theory and Practice," *Social Theory and Practice* 34, no.3 (July 2008); Somin, "Voter Ignorance and the Democratic Ideal," *Critical Review* 12, no.4 (Fall 1998).

[16] Daniel Kahneman, J.L. Knetsch, and Richard Thaler, "Anomolies: The Endowment Affect, Loss Aversion, and Status Quo Bias," *Journal of Economic Perspectives* 5 (1991); William Samuelson and Richard J. Zeckhauser, "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty* 1, no. 1 (March 1988).

status quo bias. There is thus a risk that, owing to status quo bias, non-ideal theory will understate the severity or importance of certain social problems, or even fail to see the problems as problems. The problems may instead simply be deemed "social facts" that any theorist would need to accommodate.

b. Belief in a Just World. People tend to believe that their society is just, and this increases their tendency to seek out and credit justifications for actions and events that would otherwise threaten this belief. In short, there is a tendency to believe that people get what they deserve, despite evidence to the contrary. This finding in social psychology has been firmly established in many experiments. In one experiment, subjects tended to believe that whoever won the lottery was a harder worker than a losing contestant, despite the fact that the lottery was described as a random drawing. In another experiment, subjects observed a video of participants receiving painful electric shocks. Though there was no relevant information provided to the subjects about the participants, subjects tended to develop negative opinions of the suffering participants.[17] The subjects interpreted the situation in such a way that the participants had to deserve their unpleasant experience. The phenomenon of "blaming the victim" is an upshot of belief in a just world.

It has been shown that to the extent that we have this belief in a just world, we will be less sensitive to actual injustices.[18] The avoidance of cognitive dissonance will lead people to interpret events which might have otherwise signaled injustice as deserved, and hence, just. One study looked at the correlation between the belief in a just world and one's views about disadvantaged groups. The subjects who scored high on an instrument designed to test the strength of one's belief in a just world were more likely to describe the situation of the disadvantaged groups as just.[19]

Belief in a just world, then, distorts our view of the world, and thus is a further obstacle to the accurate understanding that non-ideal theory depends upon.

c. Adaptive Preference Formation. Sometimes our overall preferences change based on what we perceive to be the available options. I may want to become an astronaut, but when I learn that because I am nearsighted I am disqualified from doing so, I may decide that being a philosopher is much better than being an astronaut, anyway. This is adaptive preference formation. As Cohen puts it, "adaptive preference formation is an irrational process in which a person comes to prefer *A* to *B* just because *A* is available and *B* is not. That *A* is more accessible than *B* is not a reason for thinking that *A* is better than *B*, but *A*'s greater availability can nevertheless cause a person to think that *A* is better."[20]

---

[17] Melvin J. Lerner and Dale T. Miller, "Just world research and the attribution process: Looking back and ahead," *Psychological Bulletin*, 85 (1978) pp. 1030-51; Melvin J. Lerner, *The Belief in a Just World: A Fundamental Delusion*, (New York: Plenum Press, 1980). Charity Scott, "Belief in a Just World: a case study in public health ethics," *Hastings Center Report* 38, no. 1, (January-February 2008). See also the discussion at http://www.scu.edu/ethics/publications/iie/v3n2/justworld.html.

[18] Zick Rubin and Letita Anne Peplau, "Who Believes in a Just World," *Journal of Social Issues,* 31, no. 3 (1975) pp. 65-89.

[19] Claudia Dalbert and Lois Yamauchi, "Belief in a Just World and Attitudes Toward Immigrants and Foreign Workers: A Cultural Comparison Between Hawaii and Germany," *Journal of Applied Social Psychology* 24, no. 18 (July 2006) pp. 1612-1626. See also Roland Benabou and Jean Tirole, "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics* 121(2): 699-746 (May, 2006).

[20] G. A. Cohen, *Self-ownership, Freedom, and Equality* (Cambridge: Cambridge University Press, 1995), p. 253. See also Jon Elster, *Sour Grapes: Studies in the Subversion of Rationality* (Cambridge: Cambridge University Press, 1983).

The status quo is more readily available than any other imagined alternative. Because of this, we may be led to rank it higher than we would were we to consider the matter in a more impartial light. The idea is similar to status quo bias, but with a different focus. With status quo bias, it is our judgment of *elements of our world* which are corrupted by the bias. Our standards remain fixed, and what changes is our view of how different elements in the world meet those standards. With adaptive preference formation, it is our judgments of what is preferable, or here, *what constitutes a just arrangement or outcome,* which are corrupted. Adaptive preference formation gets us to change our standards or ideals or exemplars.

Non-ideal theory is intended to provide us with some kind of guidance about how to make our world just. But if we are subject to adaptive preference formation, we may be mistaken about what counts as just. Our conception of justice may be corrupted by familiarity with our current arrangements, and non-ideal theory could, by directing us to focus on these arrangements, direct us away from justice.

d. Path Dependence. If we start with a picture of our non-ideal world, many of its problems will be problems of existing institutions. These problems could be solved by fixing these institutions. Yet if we concentrate our efforts on fixing these institutions, we may underestimate the value of replacing old institutions with new ones. The cost of such drastic change could be much greater than the cost of reform, in the short to medium term. With a modest enough time-horizon, it could always be more rational to fix our current institutions than to replace them. So if we start with our current institutions, and it is more cost-effective to repair them, we will be limited in the future to what we can do with those institutions. But what if justice requires a completely different set of institutions in order to be better achieved? We will be focused on the wrong thing: the institutions we are stuck with. The force of this concern depends on how strong the effects of path-dependence are and what the relative costs of repair and replacement are.

Non-ideal theory approaches justice with a picture of the real world in mind. If that picture is inaccurate, then we have reason to be skeptical about the adequacy of the resultant theory. When we look at the real world, we are subject to certain cognitive biases and reasoning errors. These lead us to mistakenly accept certain features of our world as acceptable (status quo bias, belief in a just world) or improperly adjust our normative aims (adaptive preference formation, path dependence). If our theory of justice takes existing injustices as inevitable, it is unlikely it will lead to their disappearance. Instead we'd have their continuation, which would result in relatively "bad effects." Additionally, if we do not set our normative sights properly, we will risk the persistence or development of "bad effects." The result is a distorted view of the world. It is as if we are doing political philosophy while wearing beer goggles.

Non-ideal theory, then, is subject to a version of the bad effects criticism. The bad effects will not *necessarily* follow from non-ideal theory. Whether they do will depend on our susceptibility to the four distorting factors described in this section, and perhaps other

distorting factors, as well. This is an empirical, contingent matter, but that is not a reason to dismiss it.[21]

One could ask if the four distorting factors affect *ideal* political theory, as well. While this is an open empirical question, it seems to me that the answer would have to be at least *not as much*. For when we start *not* with a picture of our current society, but with a picture of the ideal society, we are not thinking about something as likely to trigger the biased thinking.

5. Non-ideal theory and the realm of possibility

Another source of difficulty concerns our thoughts about what is possible in the future.

Farrelly asks that our political theories be "*realistic* about what the best of foreseeable conditions are."[22] Robeyns adds that we "need to take into account a wide range of feasibility constraints."[23] The concern of these and other thinkers is that if our theories of justice fail to be realistic about what is possible, they will be ineffectual at best, but likely harmful.

Can we properly identify the appropriate realm of possibility? In this section I provide some reasons to be skeptical of our ability to do so.

a. Paleo-Futurology. Paleo-futurology is the study of past predictions of the future.[24] Past visions of the future are a mixed bag of near misses, lucky guesses, inexplicable wrong turns, and cultural projections. Consider an article from 1950 predicting what life would be like in the year 2000.[25] The author, Waldemar Kaempffert, comes close with some of his predictions, particularly on technological questions; for example, that people will regularly shop by picture phone (akin to Internet shopping) and that much manufacturing will be performed by automated machines.

Where he and other futurologists tend to go wrong is in identifying the cultural background against which the fruits of technological progress are enjoyed. In the year 2000, says Kaempffert, the life of the housewife is very different than it was in 1950. He details the inventions, from plates which the housewife will "melt" down the drain, rather than wash, to the waterproof furniture that the housewife will hose off, rather than dust, to videophones, over which the housewife will inspect and order fabric. What fails to occur to Kaempffert is that the structure of the family and our attitudes about gender relations would be so radically different fifty years hence that discussions of the housewife of 2000 come off as somewhat ridiculous. Here, the futurologist fails to predict massive social change about women in society.

Sometimes futurologists predict changes that fail to come about. Kaempffert imagines the man of the house smearing his face with a depilatory lotion, instead of shaving. No man I know of does this. Why not? According to some recent research, as

---

[21] After all, not only is ideal theory's vulnerability to the bad effects criticism also an empirical, contingent matter, but it is ideal theory's failure to take certain empirical contingencies seriously that makes it vulnerable to the bad effects critique in the first place.

[22] Farrelly, "Justice in Ideal Theory," p. 853.

[23] Robeyns, "Ideal Theory," pp. 349-50.

[24] As far as I know, the term paleo-futurology is owed to Matt Novak, whose website, paleofuture.com, is a great source of information about past visions of the future.

[25] Waldemar Kaempffert, "Miracles You'll See in the Next Fifty Years," *Popular Mechanics*, February 1950.

societies develop economically and socially, the result is a *greater* display of the differences between the sexes.[26] Specifically, as societies get wealthier, males display more characteristically male traits. Depilatory lotions have long been the province of women, so its use by men would be overly feminine.

So in one sense, gender differences are eroding, insofar as women are capable of exploring options beyond the traditional housewife and mother roles. In another sense, as society gets wealthier, some gender differences are reasserting themselves. And so, the trend in male facial hair removal is not towards depilatories, but towards dangerously masculine old-fashioned razor blades.[27]

Whatever the details, the lesson is to recognize just how confusing it is to predict how things will change—particularly people's attitudes.

b. Increased Rate of Change. There is a sense that the pace of technological, economic, and social change is accelerating. I can martial no evidence for this claim, and cashing out the idea of a generic "rate of change" in any philosophically rigorous sense is beyond the scope of this paper. But I think the idea is fairly intuitive. Predicting what might happen ten years down the road is difficult. But if there is more relevant change per decade, then such prediction is even more difficult—there are simply more events and actions the effects of which one would have to trace out.

c. Knowledge Gained Exclusively Through Experience. Whether a theory of justice is worth implementing will depend, at least in part, on its effects. Some of these effects may be easier to predict than others. One particularly difficult kind of effect to predict may be what the subjects of a regime will come to think of the theory of justice implemented by the regime. Will they come to endorse this theory of justice, or develop attitudes consistent with it and which contribute to its stability? Or will the theory fail to generate support for itself, or backfire?[28]

There is some knowledge about ourselves that it seems only experience can provide. If *what we'll think about the theory of justice we're about to choose* fits into this category, then we have a problem, for some information that would be worth having in advance of the decision to implement a theory of justice is by its nature only available afterwards.

d. Conservatism in Some Social Scientific Tools. While the history of prediction gives us some reason to be skeptical of the predictive power of the social sciences, I am interested in considering a more provocative thesis—one which requires quite a bit more substantiation than I am able to present here. This thesis is that (at least some of) the standard social scientific tools of prediction are inherently conservative in a way that renders them less useful in learning about novel evaluative attitudes.

---

[26] David P. Schmitt, Anu Realo, Martin Voracek and Jüri Allik, "Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures," *Journal of Personality and Social Psychology* 94, no.1 (January 2008) pp. 168-82.

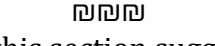
[27] One indicator: in Manhattan, men can sign up to take "Cut Throat 101" at a West Village barber shop. Seth Kugel, "Forget Shorty's Rules and It Could Get Ugly," *New York Times* (December 11, 2008), p. E3.

[28] We might continue: Are there attitudes or beliefs that are not directly about the theory but that are nonetheless relevant to the successful implementation of the theory? And if so, which of these beliefs are likely to come about?

All approaches to the social sciences rely on induction—the idea that the future will resemble the past. One thing that marks off reasonable induction from problematic conservatism is the scope of the inductive claim: *how much* of the future will resemble the past?

Consider laboratory experimentation. This kind of experimentation takes place in a controlled environment meant to capture or mimic the relevant aspects of the real world. One question about this method concerns the grounds we have for thinking that the actual future, one in which the proposed changes are implemented, is relevantly similar to the controlled environment. If we are entertaining *novel* ideas about justice or social organization, we may have grounds for doubting this similarity.

Alternatively, consider the use of models, for example, game theory. In game theory, the "players" are, in a sense, given a set of evaluative attitudes—an algorithm which governs their behavior—and their resultant behavior in various circumstances is plotted or observed. If we are to be able to learn from this, it must be because either the attitudes of the players, or the patterns of behavior we observe of such players in their controlled environment, are relevantly similar to future people. But again, it is unclear how this approach helps understand novel evaluative attitudes that may arise in response to the implementation of a new theory of justice. Game theorists can program players with a very complicated algorithm that gives the impression of an agent changing her evaluative attitudes, either in response to particular stimuli or randomly. But if the problem we are using game theory to solve is how evaluative attitudes change in response to *novel* stimuli, we are at a dead end, since, lacking exactly that knowledge, we cannot program it in advance.



The ideas briefly described in this section suggest that we may be rather poor at predicting future change. If we are bad at predicting change, then, we will have a distorted view of what is possible for us (or even, perhaps what changes would be good for us). If being *realistic* means staying within the bounds of practical possibility, we will be unable to know if we are being realistic.

If we do not know what is realistic, we are nonetheless unlikely to abstain from making judgments about the feasibility of different theories. And given our biases towards the status quo, we are likely to be cautious in counting as feasible theories which deviate from what we are accustomed to. As a result, there may be theories which would, if implemented, bring about substantial improvements on the status quo, but which we may hastily reject as unachievable because they are unfamiliar. Non-ideal theory, then, is susceptible to another version of the bad effects criticism. For any theory that endorses institutions different from the status quo, its feasibility will be unknown (while the feasibility of the status quo is known), and so we have a reason to reject the theory. But if such a theory would be better for us, then being discouraged from adopting it and being stuck with our current problems is a kind of bad effect.

Furthermore, if the bad effect is largely one of missed opportunities or forgone benefits, the idea of loss aversion (from prospect theory)—that we give much more weight

to avoiding losses than to acquiring benefits—suggests we will be especially susceptible to this kind of bad effect.[29]

Overall, non-ideal theory risks providing us with both a distorted view of reality and a distorted view of possibility, and with that bad information we risk developing normative theories that will generate bad effects.

6. Ideal Theory as a Means of Overcoming Bias

Non-ideal theory has been embraced as an alternative to ideal theory on the grounds that the former can escape the bad effects criticism leveled against the latter. I have argued that a variant of the bad effects criticism applies to non-ideal theory. This shows that non-ideal theory is far from ideal as a solution to ideal theory's problems. Yet my argument does not vindicate ideal theory. I believe that ideal theory can be improved if we learn from the mistakes of non-ideal theory. If the problem with ideal theory is that it would lead to bad effects, the solution is to find a role for it in which the production of bad effects is not as much an objection. Non-ideal theory's two failures in this regard point to two roles for ideal theory.

Non-ideal theory's first problem is that it focuses on the status quo and suffers from the biases and limitations associated with doing so. The lesson for ideal theory, then, is to not focus on the status quo. And indeed, ideal theory need not take as its starting point, "how can we make *this particular real-world society* just?" Rather, it asks, "what is justice?" or "what does justice require?" The answers to these kinds of questions may produce theories that have no hope of being implemented or followed—what Estlund calls "hopeless theories." [30] But that is alright, since their primary point is not (necessarily) to be implemented, but to give us a true theory of justice.

By directing our attention away from our present circumstances, then, ideal theory can help us in two ways. First, it moves our thinking away from a primary source of distortion and bias. Second, by contemplating questions of justice outside the context of the status quo, we can gain critical distance from it, and develop standards by which we can more clearly assess our society's advantages and shortcomings. This is a traditional role for ideal theory: to help us understand and critique our own world from an external point of view.[31] Since the theory is not necessarily to be implemented, the possibility that the theory would have bad effects *if* implemented is, at least for now, beside the point.

7. Ideal Theory as Experimental Template

Non-ideal theory's second problem is that we do not have a clear sense of what is feasible, and so we may mistakenly reject as impractical otherwise desirable theories, or reject any theories that stray too far from what we know to be practical, i.e., the status quo.

---

[29] On prospect theory see Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, 263-291 (1979).

[30] Estlund, "Utopophobia." Estlund emphasizes that feasibility is not determinant of morality. Also see Stemplowsksa.

[31] Estlund writes, "Reflection on how people and institutions should be can direct our attention and energy to determining how far realism can reach. We sometimes expect too little precisely because we have no normative standard that forces the question of whether more can be realistically expected." (8).

One way to learn whether something is practical is to try it out. And one way to conceive of ideal theory is as a suggestion about what to try. I earlier wrote that we are sometimes not in a position to know what we will think about a theory of justice until after we have had the opportunity to live it. Trying out different principles of justice could provide us with enough information from which to learn whether we are the kind of people, in the kind of circumstances, for whom such principles are appropriate. How to manage such large scale experiments is a question for another time. Note, though, that experimentation is not our only source of information, and presumably we could use more traditional methods of inquiry to narrow the range of normative options to be considered experimentally.

The key point is that this experimental role for ideal theory does not depend for its success on the avoidance of bad effects. An experiment testing whether a chemical with potential as an energy source is safe is not a failure when the experiment reveals that instead, the chemical is dangerous. If we were hoping to find the chemical safe, we may be disappointed with the result, but that is distinct from the experiment being a failure. Furthermore, to avoid disappointment, or a waste of resources, we may do what we can to limit our experiments to those we have some reason to think will yield us the result we are looking for (we will not experiment to ascertain the safety of chemicals we already know to be dangerous), and to conduct our experiments safely (we will not be reckless in our methods). This is because more than the success of the experiment—as an experiment—matters to us. But again, that other things matter to us does not make the experiment itself a failure when its result is "negative."

Similarly, if it turns out an ideal theory of justice is not appropriate for us—if it brings about bad effects—this is not a failure of the ideal theory, if ideal theory is a template for experimentation. We may be disappointed to learn that a particular theory will not work for us. We may take care to make sure we are not harmed in further experiments. We may limit the kinds of experiments we take part in, etc. But this is not because ideal theory would be undermined by the bad effects of an experiment that shows us the theory does not work for us. It is because other things besides our experiment being good as an experiment matter to us.

Additionally, there is a potentially transformative effect of experimentation that should not be overlooked. By trying out something new, one may become a different kind of person, for whom new things are now feasible. This is the principle behind any kind of training. Societies too, are capable of such progressive transformation. By trying out a theory of justice that might seem a "stretch" for us, we may come to be the kind of society for whom that theory of justice (or perhaps some other new theory) is appropriate.

8. Conclusion

Conceiving ideal theory in this way allows it to avoid the bad effects criticism. It does more than that, though.

Some defenders of ideal theory, such as Cohen and Estlund, seem content with the idea that ideal theory could have no *practical* import. Yet it may seem strange to say that a political theory could be *normative* if it lacks a "to be done-ness" about it. My strategy is to reconceive *what* is to be done—from implementation to external review and experimentation—so we can avoid the strangeness.

Reconceiving ideal theory this way also allows us recognize that some of the importance of ideal theory is still tied, if indirectly, to implementation. An ideal theory, on this view, is a *candidate* non-ideal theory. We try out different ideal theories to become clearer about what is possible for people like us in circumstances like ours, with the potential side effect of changing ourselves and our circumstances in ways that affect the practicality of these theories.

Political philosophy has been accused by professionals and students alike of being insufficiently "realistic."[32] Yet we have to be realistic about our attempts at being realistic. We should also note that demands for realism in political philosophy can extend beyond suggestions about normative *content* (the ideas theories of justice offer) to suggestions about philosophical *method* (what we are supposed to be doing with these ideas).

If we want the benefits of being realistic, it is not clear we can get them by going halfway. But what I also hope I have shown is that even on a picture of political philosophy which emphasizes realism, there is still work to be done by ideal theory.

[32] For another recent version of this criticism, see Raymond Geuss, *Philosophy and Real Politics* (Princeton: Princeton University Press, 2008).