

“Where Should We Expect Social Change in Non-Ideal Theory?”

Burke A. Hendrix

Department of Government  
Franklin & Marshall College  
232 Harris Center  
Lancaster, PA 17604 USA  
ph: (607) 351-1588  
[burke.hendrix@fandm.edu](mailto:burke.hendrix@fandm.edu)

Comments welcomed!

This essay considers the relationship between ideal theory and non-ideal theory. It begins by outlining a recent attempt by John Simmons to specify the relationship between the two clearly, noting that Simmons defends the priority of ideal theory over non-ideal theorizing. The essay then considers three different conceptions of the social barriers that currently stand in the way of a more just society, coming broadly from Mill, Marx, and Foucault. Each suggests somewhat different ways of understanding where social change is more or less likely, and the essay argues that those working in non-ideal theory should generally try to place their work in relation to these or another theory of power. The essay also notes that the Foucaultian model of power gives reasons for us as academic theorists to mistrust our own political and moral judgments. In its final section, the essay considers whether, in light of these dangers, it might be best to adopt a more limited view of the relationship between ideal and non-ideal theory, in which ideal theory retains its *logical* priority but not its *temporal* priority. In other words, the paper suggests that we may fare best when we focus first of all on reducing injustices that we recognize how to combat at the present time.

Prepared for the Annual Meeting of the Canadian Political Science Association; Concordia University, Montreal, QC; June 1-3, 2010.

## I. Ideal and Non-Ideal Theory

In a recent volume of the journal *Philosophy and Public Affairs*, John Simmons (2010) outlined one of the most developed interpretations of John Rawls' famous distinction between ideal and non-ideal theory. In this essay, I want to take Simmons' article as a jumping-off point for further exploration of the relationship between these two kinds of normative theorizing, with the intention of suggesting a somewhat different relationship between them than Simmons defends. The paper's central argument will be that, while ideal theory may have *logical* priority over non-ideal theory, it may be best if we do not attempt to give it *temporal* priority. Even if one must have an implicit theory of an ideal society to pursue normative improvements in the near term, the nature of the power relations that currently obstruct justice suggests that we should be wary of attempts to make our visions of an ideal society explicit, either to ourselves or others.

The essay will proceed in the following order. First, I will outline the relation between ideal theory and non-ideal theory as Simmons sees it, paying special attention to the logical relationship that he believes holds between ideal theory and non-ideal theories, and noting the ineliminable ambiguities that he points to where non-ideal theory is concerned. Second, I will consider the kinds of barriers that may currently stand in the way of a more ideal society. I will outline three broad conceptions of these barriers, taken roughly from Mill, Marx, and Foucault, and draw out some of its implications for non-ideal theorizing of each. In the final section of the paper, I will consider the lessons to be drawn from this comparison for conceptualizing the task of non-ideal theory. (The paper was originally intended to include a further empirical element as well, which was to involve a comparison of examples of non-ideal theorizing in regard to Canadian Aboriginal politics, but given the already significant length of the paper this section was dropped.)

## II. Simmons on Rawls and the Ideal

In his recent article, Simmons defends the relationship between ideal theory and non-ideal theory that he believes Rawls to have articulated, and attempts to specify features of the argument that Rawls himself left indeterminate. The intention of ideal theory is, roughly, to imagine the most just form of social order that can be achieved while "taking men as they are and laws as they might be" (7), to borrow the words that Rawls himself borrows from Rousseau. The central goal of ideal theory is to envision the kinds of social results that are likely to occur within societies having certain kinds of foundational organizing principles within their basic laws, when the people within those societies actually act according to those laws the largest majority of the time. As Simmons notes, the assumption of near-universal compliance is important if ideal theories are going to be tolerably precise and distinctive:

If we compare the operation of societies ordered by competing principles of justice while assuming strict compliance with those principles, the different effects we observe can reasonably be taken to be wholly the responsibility of the different ordering principles themselves. So our comparison turns out to be quite strictly a comparison of the principles of justice. If instead we try to evaluate principles in terms of how societies governed by them would operate with a "normal" amount (or a certain percentage) of noncompliance with them (supposing we can even make sense of this hypothesis), we will likely find both that our evaluations yield quite indeterminate results and that the results depend on more than simply the different ordering effects of the principles being compared (8-9).

The point of ideal theory, then, is to envision the kinds of results that would occur if certain kinds of social rules were reliably followed, to test the results usually produced by the rules themselves for the individuals who must live within them. If individuals were not reliably following those rules, we would not be able to compare various sets of rules among themselves, and would therefore find ourselves unable to reach any determinate conclusions about the outcomes produced by one set of ordering principles rather than another.

Thus ideal theory is intended to imagine “laws as they might be”, and to evaluate the moral costs and benefits that these laws might bestow upon their citizens. Because ideal theory is also concerned with “taking men as they are”, however, these structuring principles under consideration must also be consistent with basic features of human psychology. At one level, this kind of psychology will be essential in deciding the normative goals that different regimes of rules are intended to realize: if we are concerned with protecting the basic moral interests of all persons, for example, we will need to make complex judgments about the “primary goods” that must be secured if they are to live well. But these kinds of psychological presumptions are important at another level as well: viable structuring principles must be those in which very high compliance by limited human beings is in fact *possible*, because principles which cannot be upheld in this way cannot serve as rules for any actual human community. Rawls himself was particularly interested in the failures of utilitarian theories in this regard (9), but we can see many other kinds of ideal structurings that would be impossible in this way beyond the cases Rawls describes. Indeed, we can see these basic psychological expectations at work in Rawls’ presumption that some form of state will be necessary in any form of human community. If principles of justice were enough to reliably secure their own adherence, law itself would not be necessary; we would all live together in peace and security without any need for the more coercive aspects of human social organization.

Ideal theory thus involves two forms of speculation: the first about the outcomes created by the operation of certain principles of social order, and the second about the possibility of compliance once predictable human failings are taken into account. It may be obvious that these two concerns are related. The outcomes of institutional rules will be determinate only insofar as they find reliable attachment points within tolerably fixed elements of human psychology, and humans will be likely to comply with these rules only when these rules trip the right kinds of psychological levers to bring such compliance about. Thus exercises of ideal theory are in many ways attempts to spell out “men as they are” – to specify as carefully as possible the features of human psychology that are sufficient for motivation, and the degree and nature of the flexibility in our inducible patterns of behavior. For Rawls, the correct suppositions about human nature suggest that a well-ordered society treating all persons as equals will be sufficient to generate its own long-term support, so that the “realistic utopia” toward which we should strive will be more satisfying morally than the world in which we currently live. Not all sets of suppositions about human psychology will lead to such pleasant conclusions, of course. One is reminded of Rousseau’s own words in the *Discourse on Inequality*, that “the flaws which make social institutions necessary are the same as make the abuse of them unavoidable” – a presumption that might lead to a very different kind of ideal than that outlined by Rawls. The specific contours of the best ideal theory are thus pretty uncertain; the overarching considerations relevant to such a theory, however, seem at least in principle specifiable.

What, then, is the precise relationship between ideal theory and specific non-ideal theories? Non-ideal theory is concerned with the mechanisms by which real societies might be brought into compliance with the terms of ideal theory, and here, as Simmons notes, there are

two kinds of potential roadblocks that are indicated in Rawls' work. The first comes from natural or historical contingencies that are difficult to overcome even if all parties are correctly motivated to pursue justice (13-15). Perhaps the easiest cases here involve severe poverty, where a society simply lacks the resources to guarantee everyone a decent standard of living while also securing for everyone the other kinds of protections (equal rights and so on) that they would deserve in a more prosperous society. Here, the ideal society practicable by people under better circumstances is not practicable by them for reasons that are not under the control of anyone's choices. Perhaps surprisingly, political and cultural traditions also fall within this category (15), even though these seem at least potentially under human control. Rawls' basic presumption seems to be that individuals are not really in control of the valuations and practices they inherit, even if they have the capacity to reform these over term in conjunction with many other persons. One might include ignorance of relevant sorts within this category as well – those who do not know what the structure of the “realistic utopia” looks like should not be held responsible for their failure to live up to its standards, nor should those who choose a mistaken path to pursuing it, and so on.

The second kind of roadblock to compliance with the standards necessary for the realization of ideal theory come from those who deliberately choose to diverge from those standards even while recognizing their moral force (15-16). While they may do this simply at an individual level, they may also do this in some collective way, when multiple self-interested actors use their social leverage to block important kinds of social changes in protection of their own current or expected advantages. As may be obvious, patterns of injustice within actually existing societies are likely to be a combination of both elements. Many of the behaviors that we might think as intentional blockages of justice are likely to result in part from persons believing that they are entitled to behave in this way (e.g. on the belief that market society or pluralist democracy permit them to pursue their own interests through any currently legal means), and in many cases opportunistic manipulations of existing regimes of social power will be possible only because of such mistaken notions among large numbers of people. Thus it is probably best to see both of these kinds of blockages working together within actually existing societies, rather than operating in distinctive aspects of human societies (cf. 15-17).

In either case, the challenge for non-ideal theory is the same: to find ways to move persons and their circumstances toward the set of structural arrangements characteristic of the ideal society. Ideal theory is thus logically prior to non-ideal theory (34), and sets the possible ends toward which it can strive. Stated in this general way, the relationship seems relatively clear. Yet it just as obviously leaves a great deal of uncertainty about how one should move toward the ideal from non-ideal circumstances, and if we are really interested with moving in this direction, it seems incumbent to say something considerably more detailed. Simmons, drawing from Rawls, offers a number of standards for the political choices of actors under non-ideal circumstances (18-19). First, they must make choices that are (A) morally permissible, in the sense that they do not too profoundly infringe the legitimate moral interests of anyone (something which neither Simmons nor Rawls defines with much precision). Second, they must be (B) politically possible, in the sense that there is some hope that they will actually be able to be put into place, given political circumstances as they are. Third, they must be (C) likely to be effective in actually bringing about movement in the correct direction, and fourth, they must be structured to (D) reduce more grievous injustices before less grievous ones.

According to Simmons, the nature of ideal theory suggests that strategies for non-ideal circumstances ought to be directed as strongly as possible toward the overall goal of a just

society. When we have questions about which out of the range of politically possible strategies to pursue, we should choose the one that has the most promise of leading us as directly as possible to the correct overall organization of society, rather than to something more fragmented:

Because the object of Rawls' nonideal theory is the eventual achievement of the ideal of perfect justice, not simply the elimination of particular or salient injustices, I take nonideal theory's requirements of political possibility and likely effectiveness to be best understood...as requirements that policies be politically possible and likely to be effective as parts of a strategy for the complete elimination of all society injustices. This means that we must understand Rawlsian nonideal theory as both strongly transitional (as opposed to simply comparative) in character and as offering us an integrated, rather than piecemeal, goal as our target (for assessing policies' possibility and effectiveness) (21-2).

Thus, Simmons argues, non-ideal theory should not be concerned primarily with ranking the moral status of any two social options directly against one another without reference to this longer-term goal (25), but instead with ranking them in terms of their long-term contribution to the goal of a just society as well. In some cases, this may mean pursuing morally worse short-term policies, when these allow for paths forward that other policies would permanently foreclose (25).

If we find everything that Simmons has argued for plausible, it seems clear that massive gulf still remains in thinking about what these kinds of evaluative tools may commit us to in any specific circumstances. Simmons admits the complexity of the considerations that will go into making even a single recommendation for a real society in this regard:

Although much is obviously left vague here, we can at least infer this much...: nonideal theory will require judgments of both philosophical and social-scientific sorts. Determinations of a policy's "moral permissibility" obviously lie in the proper domain of moral and political philosophy, as do judgments of grievousness, which depend on prior ideal theorizing. Determinations of "political possibility" and "likely effectiveness," on the other hand, seem more naturally to require the expertise of, e.g., political scientists, economists, and psychologists. Further, since the general principles of nonideal theory are intended to have implications for the policies of particular societies, all three kinds of determinations will require reasonably specialized knowledge of the structure and workings of these particular societies (19).

Effective instances of non-ideal theory will thus require the cooperation of moral philosophers and social and behavioral scientists in ways that do not always happen naturally in academic or other practice, and seem likely to often require the cooperation of social actors on the ground as well, particularly since they are likely to have far more of the specific knowledge of social and political possibility that others lack. But the general structure of how we should proceed is clear enough: clarify the ideal to some tolerable level, decide what the sources of impediment to it might be, and put in place strategies to overcome them.

### III. Where Should We Expect Moral Failure?

It would obviously be easier to proceed with something more detailed if we had a general theory of moral failure. That is, what kinds of obstacles to an ideal society are likely to be most resistant to reform, and which the least? While such a theory seems inevitably hard to develop, it seems mistaken to believe that these kinds of judgments can only be made locally, at the level of

any specific society at any specific point in time. Rather, human societies seem likely to have recurrent patterns within them that make change in some directions almost impossible, and make change in others quite easy. Having a clearer view of these general patterns seems essential if we are to think about the nature of non-ideal theory in more specific terms, however complex the issues involved.

While moral philosophers are generally nonspecific about the nature of social power within unjust societies, the very nature of their project and of the kinds of conclusions they reach suggests that they see the most dangerous countervailing forces to be (i) ignorance of empirical facts, (ii) poor reasoning, and (iii) self-interest, without any detailed specification of the relationship between them. Detailed efforts to outline the nature of an ideal society suggest that most philosophers believe the resolution of factual errors and logical inconsistencies to be central to the social task of creating a more just society. Once the right kind of theory and the necessary kinds of knowledge are available, morally motivated people will be able to move society in the appropriate direction, with inevitable difficulties along the way from those who remain in ignorance or dominated by their narrow interests. Once we have achieved this ideal society, the assumption usually is it will be stable for the right reasons, because it balances or otherwise overcomes persistent human weaknesses in the maximally effective and appropriate way – this is, after all, the very definition of an ideal society.

This usual method of proceeding assumes substantial reliability in our processes of moral thinking and methods for recognizing details of the empirical world when viewed over the long term. While few moral philosophers would expect specific individuals to do exceptionally well in this regard on their own, they do generally expect that a community of free inquirers will approach reliability over the long term. Allen Buchanan (2004, 2002), for example, has argued plausibly that many of the institutional features associated with liberal societies can help to maximize the opportunities and motivations that individuals have to seek correct information, and to hold one another accountable for their informational errors. This is an old argument within liberal theory in one form or another. Buchanan (2002, 141-4) links his own version to that of Mill, and there are many other versions that reach similar conclusions. All assume that liberal institutions and especially freedom of speech will tend to strengthen both themselves and the cause of justice over time, with recursive benefits that will continually bring us nearer to where we hope to be. Restated, liberal institutions that encourage the free circulation of information will help to disassemble the mistaken ideas that stand in the way of justice, so that it can be both recognized and achieved.

While this is a compelling image, it does not on its own tell us where the most obdurate sources of resistance to long-term justice are to be found. Presumably there are certain kinds of mistaken ideas that are more resistant to change than others, and recognizing which ones these are seems essential if the process of non-ideal theorizing is to go forward effectively. Certainly it will be very hard to pursue any unified program of non-ideal theory of the kind that Simmons suggests in the absence of such an understanding, because we will have difficulties in predicting whether specific social changes will lead more or less swiftly to the long-term goal (or even at all). We may face more localized problems as well: if we cannot make reliable predictions about where patterns of injustice are robust rather than vulnerable, it may be difficult to anticipate harms to many legitimate moral interests from choices made by particular social actors even without regard to overall patterns.

To proceed most effectively, then, ideal theory should be supplemented by a more general theory of power. Since most of us are concerned about justice within our own imperfect,

quasi-democratic societies, it makes sense to seek to explain large-scale and robust patterns of obstacles to justice among members of societies such as our own. As a rough beginning, three broad kinds of options seem available for conceptualizing the barriers to a more ideal world. Although each is overly simplistic on its own, it seems best to lay them out in relatively stark form. The first potential explanation, associated most strongly with Mill and hinted at above, views the primary blockades to justice as intellectual, primarily in the form of historical ideas held over in the absence of sufficient moral reflection. Following Mill, we might call this the problem of custom. The second, associated most strongly with Marx, regards problematic patterns of belief as mostly epiphenomenal expressions of more basic structural arrangements in the economic and productive sphere. On this view, the primary obstacles to justice are not primarily intellectual, but material and organizational. The third possibility is that associated with Foucault, which sees social structures and conceptions of human knowledge as mutually imbricated in complex but usually specifiable ways. Here, the problem is that particular conjunctions of material and intellectual life are tightly bound in self-replicating ways, usually through a large and complex set of (sometimes competing) social institutions.

In Mill's account, the primary obstacle to justice lies in intellectual inheritances from previous eras, which continue into the present primarily through a lack of critical evaluation. From this beginning point, there seem to be two broad and intuitive criteria for evaluating the potential of particular courses of social change. The first is depth of agreement on a particular custom: all things considered, the more broadly an opinion is held, the more difficult it will be to change. This is potentially counterbalanced by the second kind of consideration, however, which is in some ways an outgrowth of the first: the more logically inconsistent a belief is with other broadly-held beliefs, the more liable it is to change. This suggests a fairly intuitive principle for non-ideal theory. Generally, exercises of non-ideal theorizing will proceed most effectively when they focus first of all on "vulnerable" beliefs – those that seem to be held by less than a majority, or that are wildly out of balance with broader social understandings. The most effective exercises of non-ideal social action are likely to be those that spend less time discussing some distant ideal, on this interpretation, and more on exposing some particular logical flaw in a sharp and incisive way. This obviously leaves much yet to be resolved, but it begins to give us a sense of relative possibilities, which must then be coupled with normative judgments about relative grievousness, moral permissibility, and so on.

A roughly Millian conception of the obstacles to social change seems to be the one most commonly adopted by moral theorists. Because the core features of this conception are usually unstated, however, theorists often find themselves falling into different conceptions of their own task in moral theorizing, with resulting tensions that have been obvious in many of the debates surrounding the relationship between ideal and non-ideal theory. If the exposure of vulnerable beliefs is to be our goal, we may often wish to engage in relatively simplified exercises in argumentation, which seek to engage with large numbers of people in relatively stark terms. If we also wish to develop an ideal theory conception of social possibilities, however, we may be required to hold both an "inner" and "outer" theory, one to guide our own thinking and one for public use. This is not an impossible balance on its own terms. Despite our aspirations, most of what we write in the academy can only be for "inner" use, because few outsiders will have any interest in it. But it is nonetheless a difficult balance to strike well, and many of the vehement battles that have gone on over the relative social value of ideal theory have involved questions about whether to maintain this division, and if so where. On this kind of model, there is likely to remain a continual split between theories that are comprehensive and logically sound and those

that are social winners. Unfortunately, this split can weaken both when it comes into the open: when revealed, it suggests to the general public that the defensibility of their principles is a matter of debate, and can even leave theorists perplexed about just what their project is. Similarly, differing judgments about grievousness or achievability can lead to battles about where to place one's attention, and similar weaknesses of impact. One might hope, with Mill, that over the long run the free exchange of information will lead to the best overall outcomes, but it is not of much help for any of our actual and specific choices, and it is this that non-ideal theory is intended to achieve. If one holds this conception of the impediments to justice, then, it seems that we will fare best when we are relatively systematic about specifying who our intellectual efforts are intended to target, and about how we hope to keep appropriately distributed to their intended audience. (Density of writing and deployment of jargon is of course a familiar tactic for managing such distributions.)

Marx's alternative conception of power can be stultifying if we take it to foreclose all human agency entirely, but there is no reason to adopt something so stringent. Rather, we can see the core argument of the Marxian account to be that social change is easy in ways that are in accord with contemporary economic patterns, and difficult in directions that clash with them. It would likely be a mistake to see the Marxian account of power as primarily focusing on the economic self-interest of individuals. Rather, it focuses on the structural needs of systems of material production, in ways that involve individual self-interest in a somewhat subsidiary role. If it is the case that our means of producing material goods are also our means of producing ourselves as persons with particular characteristics, then it seems that we should focus first of all on the ways in which economic systems broadly construed structure human society in their own image. Human beliefs about self-interest, on this account, can themselves be shaped in important ways by background mechanisms of material production.

While it does not seem helpful to become bogged down in questions about the precise scope of agency here – clearly it would be unhelpful to imagine it away entirely – the intuitive implications for social change within this conception seem to be plausible ones. While the Millian model suggests that widely-held beliefs are likely to be the most resistant to change, the Marxian model suggests that such beliefs may be relatively easy to change if they are out of phase with the requirements of economic production. (Thus Marx argued that the increase in the rights of women in his day was related to increased need for laborers, for example.) Similarly, intellectual frameworks will be difficult to change if they are necessarily associated with structures of production, even if many of their elements are mutually contradictory with one another. Historical continuity is thus no guarantee that beliefs will continue to exist, nor is incoherence necessarily a useful lever for reformers. In calling upon us to reconsider arguments about custom, the Marxian account sounds important warnings: long-established ideas can sometimes change rapidly when social circumstances do, even if no one exactly intends this or even recognizes it, and the revelation of even severe intellectual incoherencies is no guarantee of even long-term success. If the views of most ideal theorists are roughly correct in economic matters, the continuing resistance to these ideas would seem to be ample illustration of such forces.

If we adopt this conception of the obstacles to justice, the most successful interventions by moral theorists are likely to be those that exploit instances in which morally problematic beliefs or practices have become out of phase with forces of material production. Since these kinds of instances will always be limited and imperfect so long as social change remains driven by material forces substantially outside human control, it seems that the task of non-ideal theory



may become a permanent one: it will be continually necessary to try to make morally beneficial improvements since many may be reversed as circumstances change. This need not be the case, of course. There may be ways to design an escape from these circumstances with small nudges here and there, if for example ethically-sensitive inventors and business planners were to adopt innovations that are both consistent with existing patterns of production and likely to be produce better moral outcomes. But this account of power makes the task of non-ideal theory a daunting one even if there are routes to permanent improvement within it: one must somehow try to anticipate vulnerabilities that may not yet exist, or find ways to actively create them. As a small consolation, this account of power at least suggests that there may be limited tensions between exercises of ideal theory and non-ideal theory: both may be able to provide tools to help on those rare occasions when interventions might be most effective, and allowing as many resources as possible to exist for such agents may be the best we can hope to achieve. Insofar as beliefs are less responsible for injustices, we at least have to worry less about mixed public messages.

Many of the general warnings about the difficulty of pursuing visions of ideal theory are echoed and expanded on by Foucault's more complex account of power. While Foucault's view is never articulated as clearly as one might hope, its basic outlines seem tolerably clear. At its core, Foucault's view of power entails a deep and systematic skepticism about human cognitive capabilities that has overlaps with the Marxian view. Most obviously, Foucault argues that human intellectual frameworks are rarely causally free of structural conditioning. Because we are biological creatures, our thinking processes are highly dependent upon the conditions of our bodies; because our bodies can be manipulated, controlled, and habituated, Foucault argues, our thought processes, identities, and very experiences of life are subject to being structured in finely detailed ways by a variety of organizational methods. Foucault notes the way in which a variety of institutions, from schools and prisons to businesses, can condition us to be specific kinds of people with specific kinds of skills, capacities, and expectations, without us ever recognizing what is occurring.

Because these institutions often require the active cooperation of those who work within them, Foucault argues that institutions of power generally produce structures of knowledge alongside them. Schools produces knowledges about how students can be made teachable and manageable, for example, while businesses produce knowledge of how workers may be made productive and about the capacities that they may be able to bring to bear for specific purposes. Although there is nothing nefarious about these kinds of knowledge in abstract terms, Foucault argues that institutions will tend to survive and reproduce themselves most reliably when they most effectively direct persons toward the maximal expansion of their organizational control. Thus, in a process that is roughly akin to the "survival of the fittest" seen among firms in a market or animals in the wild, organizations will continually evolve toward greater and greater shaping and mobilization of the capacities of those ensconced within them, adopting the newest techniques of control under threat of obsolescence and disappearance. Importantly, organizational structures do not replicate themselves because of their larger social benefits, or because of their humane qualities to those who live within them; rather, they do so on the basis of their capacity to effectively shape behavior in the direction of institutional continuance and continued control. Sometimes these institutions will incidentally create public or individual moral gains, but this is not their driving force; their imperatives are rather internal and substantially autonomic.

It is hard to capture the nuance of this model of power without reference to Foucault's copious examples, and like the Marxian model this view is easy to caricature for its limited view of human agency. However, its basic conceptualization of contemporary human social life

seems to have substantial force. It clearly is the case that the contemporary world is one of multiple controlling institutions, often with systems of knowledge and enculturation of their own, and there is no obvious reason to expect that these institutions reliably point in the direction of justice, especially within a society that is already agreed to fall considerably short of what ideal theory demands. Since we are cognitively limited, inherently social, and highly teachable creatures, Foucault's model seems a plausible fit with basic features of human psychology as well. The question then becomes how we might use this model to think about non-ideal theory. This proves somewhat more complex than one might first assume, in ways that seem to me important.

On this model of power, the obstacles to social change are likely to be multiple and complex, but they are likely to point to conclusions that are partway between the Millian and the Marxian model, while leaning more heavily in the direction of the second. If it is the case that structures of knowledge are interpenetrated by structures of power and vice versa, then it seems plausible to say that many of the obstacles to justice lie in the beliefs of persons within a society as the Millian account suggests. These beliefs will not necessarily have a long history in their specific form, but they will have genealogies that may be traceable, and this suggests some overlaps with the Millian conception as well. Yet importantly, the Foucaultian account suggests that these beliefs are not autonomous drivers of human behavior; they are rather tied up within organizational and disciplinary systems that help to shape persons for whom those beliefs seem appropriate. This suggests that theoretical interventions will often have to share features with interventions under the Marxian conception of power. If it is the case that institutions tend to produce particular kinds of knowledge that help them to replicate themselves, then it seems unlikely that simply showing apparent inconsistencies within beliefs will be sufficient – often systems of belief will have appropriate flexibility or defusing mechanisms built into them already, and they will likely have multiple mechanisms to dampen certain cognitive avenues for those within them. If this is true, then like the Marxian account it may be most reasonable to expect social change along channels that are easily accessible from those the organization is already travelling.

Given the complexity of multiple competing institutions, attempted interventions may have to be carefully targeted to different systems of knowledge and organizational structure, often with the hope of pointing them in one direction or another where options present themselves. If there are multiple institutions, multiple strategies may be necessary, often shifting and diverging depending on context. On this model, we as theorists should be wary of a common tendency among normative thinkers since time immemorial: to think that we only need to capture one central institution, usually the state. There are multiple problems here. First, we are not going to capture the state in any case, any more than philosophers were about to become kings in ancient Athens. Second, and perhaps more importantly, “the state” is not a single organization with a single organizing principle: it is instead a network of complex administrative, coercive, and other structures, and these do not always respond in the same way to attempted alterations. Nor can we easily say that “the law” will require these individual elements to behave appropriately, if it is just once structured in a particular way. The law, insofar as it is a unitary phenomenon, is itself a system of knowledge that exists among a particular set of human beings caught within complex organizational systems, and it is not more free of its own internal mandates than any such system. Systems of law can achieve some things, but not others, and not always in the ways that one might hope or imagine. Certainly the law is a tool of limited value

so long as competing forms of organizing persons continue to exist, as they surely must in any kind of human society that is reasonably complex and materially productive.

This suggests two broad possibilities for ideal theory. One is to try to seek for a single, overarching strategy that will create a unified program for pursuing social change across all of these multiple venues. On this standard, one would try to directly pursue an overall program of ensuring that disciplinary institutions become sufficiently restrained behave in ways that can, in some loose sense, be approved by those who live within them. (This is broadly the point of Rawls' theory, in many ways, given its concern with the multiple forces that can shape individuals in both their capacities and preferences.) An overarching approach would probably continue to look in part toward methods of strengthening the capacities to law to regulate aspects of society that current elude it, and then would seek for means to ensure that law actually behaved as it was intended, so that it came to most closely approximate our hopes for it over the long term. But it would also look for ways in which one could intervene in specific kinds of disciplinary institutions, while trying to chart a broad patterns of interventions that might have the desired outcomes if they all came together correctly. The second model, on the other hand, would be more tactical and piecemeal, without any attempt to chart a master strategy to the correct long-term outcome. Rather, it would try to diminish specific kinds of injustices located within particular institutions first, without focusing on the longer term question of how to actually reach our ideal theory goal.

Perhaps counterintuitively, the conception of power found in the Foucaultian model suggests that we may often be better off adopting the second option rather than the first, if we are concerned about maximizing our approximation of justice over the long term. The problem is this: a detailed plan for how to intervene in society over the long term would itself require the cooperation of a large number of persons working through institutional means and across networks of communications, and the Foucaultian model of power suggests that these kinds of mechanisms are likely to begin producing their own effects in structuring the beliefs of those who live within them. Such systems will begin to form kinds of knowledge that are specifically important to them, and that are necessary for anyone who seeks to navigate them effectively. Insofar as non-ideal theory requires the cooperation of moral philosophers, psychologists, social scientists, and others, it will tend to gravitate toward particular methodologies for approving or disapproving of beliefs, and toward particular kinds of common terminologies, research questions, and recurrent questions. If Foucault is correct, these systems of knowledge will be at best limited and partial, with their own peculiar mandates and obsessions that will have more to do with patterns of coordinating and overseeing members than with the pursuit of truth and morality as such.

In other words, the Foucaultian model sounds a note of warning that is mostly absent in the Millian and Marxian models of power: it calls into question our own professional judgment about how to create moral improvements over time. For the Millian model, the primary problem is custom and the weight of history, and the best remedy is a careful testing of ideas for factual accuracy and coherence – skills at which academic philosophers have reason to believe themselves especially competent. For the Marxian model, the primary source of belief lies in the economic imperatives of productive forces as a whole. When we as theorists call into question the patterns of distribution created by those productive forces, it is a natural temptation to believe that we have broken free of the most substantial social forces fostering false belief. In either case, we are likely to regard our own reasoning as relatively untainted by broader social conditions. For the Foucaultian model, on the other hand, we should be especially wary of

trusting our own expectations: systems of knowledge are strongly correlated with the continued replication of specific forms of social organization, and the academy is manifestly a social institution with multiple mechanisms of shaping and disciplining its members and the knowledge they produce. Some of the mechanisms here are crude and obvious. As academics we generally receive employment, raises, and social acclamation based on the size of our academic audience, and not on our actual contributions to a world more resembling that hoped for by ideal theory. But there are more subtle problems as well. In teaching a relatively unified canon of important thinkers, for example, we not only signal that certain kinds of ideas are more important than others, but we simultaneously produce social actors for the future who have learned to think from within these particular viewpoints, whether they ultimately prefer to do so or not.

The broad conclusion of this model of power is that we have reasons to distrust our own reasoning, our speculations about what the nature of an ideal society might actually look like. If ideal theory is concerned with “men as they are, and laws as they might be”, it seems reasonable to expect that we will be unreliable judges of both. We may be unreliable as judges of what kinds of creatures we as humans are – what kinds of needs we have, what kinds of motivations drive us, what kinds of weaknesses we can overcome – and also of how systems of law and social organization might actually function in practice. On this model of power, there is an extra and centrally important blockade to the realization of an ideal society: ourselves, and our incapacities in reliably specifying what an ideal society might look like. If this is the case, it might be that the Foucaultian model of power suggests something far less rigid and systematic if we are to actually pursue a more ideal future. At a minimum, it might be best if those working toward an ideal society *did not* try to coordinate their efforts, and did not try to specify too carefully the path that they hoped for society to tread.

#### IV. Achieving Justice Or Reducing Injustice?

Endorsing one model of power rather than another makes it likely that exercises of non-ideal theory will vary widely. Given these divergences, not only in the specific recommendations to be made but in their differing conceptions of the task of non-ideal theory itself, it seems important for those who wish to engage in extended exercises of non-ideal theory to specify as clearly as they are able the conception of power with which they are working. This is important not only to allow others to understand the nature of the argument quickly, but also to ensure consistency within and across theoretical writings themselves. It is always difficult to decide which features of the social world to regard as relatively resistant to change and which as vulnerable. The very task of normative theorizing itself presumes that moral argument can make a difference, and this makes it hard to decide where to cease theorizing and to simply take features of the social world as fixed for the purposes of moral exhortation and analysis. If we are not systematic about our conceptions of power, there is a good chance that we will unknowingly change our expectations from case to case, so that we assemble conceptions of social change that are far more plausible in some of their recommendations than in others. If non-ideal theorizing becomes a plausible task only when we can reasonably anticipate the outcomes of various changes in conjunction with one another, this poses obvious difficulties. This is a permanent problem of non-ideal theorizing in any case, given our persistent lack of knowledge, but that is a reason to try to reduce the dangers, rather than to ignore them.

The Millean, Marxian, and Foucaultian models of power differ primarily in their conceptions of how non-ideal theory might hope to be most persuasive toward its potential targets, and in the degree to which they expect past intellectual conceptions to be determinative

of current obstacles to justice. For the Millian model, the past is the best predictor, because it is the source of the dangers itself. For Marx, the past is only a weak predictor in cases of technological or economic change. Although one can predict that economic forces will drive human life in one direction or another, one cannot often recognize what that direction might be in advance. The Foucaultian model, unsurprisingly, seems to fall somewhere in between: past social structures that organize power/knowledge in particular ways are likely to continue to persist, but they may sometimes change in unexpected directions as well. The Foucaultian model also suggests the aforementioned reasons to believe that we may be substantially unreliable as evaluators of the empirical and moral order of the world, because we are ourselves participants in regimes of power that channel our thought in particular directions.

Which of these models is the correct one? Answering this question would require entering some complex territory, including epistemology, psychology, linguistics, and so on. Yet one should be extremely wary of proceeding with an exercise of even ideal theorizing until one is confident of what these social forces might be. If ideal theory is intended to take “men as they are” and “laws as they might be”, then we obviously should be wary of saying too much about laws before we come to a full understanding of men. If we believe that an ideal society must be “stable for the right reasons”, for example, then we will be unable to determine which sets of possible arrangements will be stable until this initial work is completed. It may be the case that we cannot plot out the details of an ideal society until we have already determined in large measure the set of social changes that are possible in contemporary circumstances. It may be the case that this cannot be determined mostly in advance; at least potentially, this will have to be worked out as we go.

On the Millian and Marxian models of power, there may not be reasons to forgo ideal theorizing in the meantime, even if they are unlikely to be successful. At worst, they may slow down the process of moving toward justice somewhat. (This process is likely to go at radically different speeds within each model, of course.) On the Foucaultian model, there seems to be greater cause for concern. If it is the case that we as members of the academy are ensconced within social institutions that play a fundamental role in shaping ourselves as members and generating future forms of social discipline, then there may be especially important reasons to be wary of speculations about a future ideal world. If our ideas are likely to be channelled in particular directions by our social circumstances, and those ideas themselves are likely to serve as channels for future kinds of organizational power, then premature attempts at ideal theorizing are themselves likely to be dangerous and morally counterproductive. They may actually help to replicate or maintain injustice, rather than to combat it, even if we understand our own actions in precisely opposite ways. It would not be hard to find historical examples in which extraordinarily thoughtful and well-meaning people have behaved in exactly this way. The question is whether we have reason to believe that we have managed to escape from those kinds of limitations, and why.

My own sense is that we cannot yet specify with much certainty the social forces that stand in the way of movement toward a more ideal society in the present day, so that we cannot easily make determinate statements about the “men” side of the men and laws equation. (I doubt we can say much about laws that is especially revealing, either.) In these circumstances, there are reasons to believe that we may pursue the goals of ideal theory more effectively if we do not try to pursue them directly. If it is always very hard to predict where social change is possible, then it may make sense not to begin planning too heavily for multiple interlinked changes. Moreover, it likely makes sense for us as normative theorists to pay more attention to what, for

example, activists may be saying on the streets, in their meetings, and in their newsletters. People who are directly attempting to foster social change seem likely to have substantial insight into where weaknesses exist in current regimes of power, and it may be that realistic ideal theories can only be worked out in small pieces as one goes forward. Given the difficulties of recognizing humans as we are, it may be the case that ideal theory will have to emerge only once we begin to see what is socially possible in the near future.

This suggests that ideal theorizing may not have *temporal* priority over non-ideal theorizing, even if it has *logical* priority. Simmons is surely right to note that, as a conceptual matter, non-ideal theory only has meaning as a method to get somewhere in particular. In this sense, ideal theory is logically prior. This does not mean that it should be temporally prior in our own actions, however. We can see examples of this kind of divergence in the natural sciences. A unified field theory necessarily has logical priority over individual explanations of gravity, the nuclear forces, and other more specific laws of the physical world. But no one knows how to begin creating a unified field theory at the moment, because there is not yet sufficient evidence of how its components might work or relate to one another to make the task fruitful. Thus we as a society do not invest a great deal of resources into direct efforts to recognize this unified theory, because we recognize it as premature. Something similar may hold with ideal theory in normative terms. At best, resources invested in ideal theorizing may be resources that are wasted as premature. At worst – if the Foucaultian model is correct – these investments might help to play a role in actually slowing the approach of a more just society.

Indeed, it is worth considering whether we should really prepare for *ever* achieving an ideal society, given the widely acknowledged limits of human cognitive, motivational, and organizational powers. Amartya Sen has argued that non-ideal theory should focus first of all on showing whether one kind of social arrangement is more relatively just than another. In imagistic terms, he suggests that one does not need to know whether Mount Everest is the tallest peak in the world to know whether it is taller than those that surround it. Simmons deploys this image in turn to argue for the logical priority of ideal theory:

Which of two smaller “peaks” of justice is the higher (or more just) is a judgment that matters conclusively only if they are both equally feasible paths to the highest peak of perfect justice. And in order to endorse a route to the highest peak, we certainly do need to know which one that highest peak is. Perhaps for a while we can just aim ourselves in the general direction of the Himalayas, adjusting our paths more finely – between Everest and K2, say – only when we arrive in India. But we need to know a great deal more about where to find the serious candidates for the highest peak before we can endorse any path from them to here (35).

The image is a powerful one. Even if we are not sure which mountain is the highest, it at least makes sense to look for the areas of the world in which the highest mountain might be found. Yet we also have to consider whether we can ever actually get to India, so to speak. If it is the case that moving between geographical regions is extremely difficult, and it is important to be on the highest peak available, rather than in one of the valleys, then it may make sense to move only from on logically recognizable peak to the next, continually hoping that taller ones are yet to be found. This counter-image on its own is of course somewhat unmotivated, but given the suffering that comes from injustice (indeed, that *makes it* injustice), it may be helpful to imagine that suffering lurks in the valleys, and that it diminishes as we rise higher. In those circumstances, it matters a great deal whether we can reach the next mountain when we set off for it, and we might hope for some brave explorers who have made the attempt themselves first

to show us that it can be reached. We would probably want to be especially wary of those who tell us it is achievable after simply looking from the heights of our current mountain itself.

At the risk of overplaying a somewhat tendentious metaphor, there is another important possibility here that is suggested by both the Foucaultian and the Marxian models. This is the possibility that the highest mountain is only sometimes in India, and other times in different locations of the world. What might this mean? If many aspects of human suffering come from very specific forms of domination or deprivation, and these forms of domination or deprivation may change over time in substantial ways as technological change occurs or structures of social control shift, it may be the case that there is no determinate specification of the set of laws that will best protect persons from injustice over the long term. It may be the case, even if we can determine the basic fixed features of human psychology and so on, that the organizational and other arrangements giving them salience may continually shift. This seems clearest with technologies: some patterns of distribution that may be sufficiently just and stable for the right reasons in one technological circumstance may become unjust or unstable in the next. Do we believe that human technological change can be stopped, if it was simply shown to be dangerous to justice? Here is something of a test for one's expectations about the relative force of the Millian model compared to the Marxian or Foucaultian. My own suspicion is that such changes simply cannot be stopped, whatever we as normative theorists might advocate. In these circumstances, then, it may be the case that Everest will sometimes move from one country to another without warning, so that our previous efforts at reaching the highest peak will become either wasted or counterproductive.

My own preference is to endorse a model of power closer to the Foucaultian model than the Millian model, and to thereby adopt a more limited conception of how non-ideal theory should proceed. My own vision of a workable non-ideal theory, at the present, is a substantially chastened one, that focuses first of all on evaluating achievable moral improvements in the cases of specific social actors. Because my area of empirical study has primarily been in indigenous politics, this has led me to focus specifically on political, legal, and other changes that might reduce avoidable injustices in regard to these specific members of society (Hendrix forthcoming). This is admittedly partisan work in many cases, since it involves the advocacy of specific political and social changes for a particular and determinate set of persons and conditions. Given the aspirations of ideal theorists, such work can often look narrow and peculiar, because it attempts to play a role in contributing to plotting one particular course for a relatively small number of people to slightly higher mountain peaks. It is understandable that this strategy might be seen as excessively deflationary. Yet political theorists are rarely the drivers of positive social change themselves. My own view is that political theorists may serve best as adjuncts to those who are. I do not expect others to necessarily share this view, of course. But it does seem important to ask precisely what it is that we as political theorists hope to achieve, and how we hope to achieve it. The relationship between ideal theory and non-ideal theory that Simmons outlines is a very attractive one. Without a stronger conception of social power, however, we as moral theorists may be disabled from saying whether it can ever be anything more.

Works Cited

- Buchanan, Allen. 2002. "Social Moral Epistemology." *Social Philosophy and Policy* 19: 126-152.
- Buchanan, Allen. 2004. "Political Liberalism and Social Epistemology." *Philosophy and Public Affairs* 32: 95-130.
- Hendrix, Burke A. Forthcoming. "Political Theorists as Dangerous Social Actors." *Critical Review of Social and Political Philosophy*.
- Simmons, A. John. 2010. "Ideal and Non-Ideal Theory." *Philosophy and Public Affairs* 38: 5-36.