The Journal of Political Philosophy: Volume 20, Number 1, 2012, pp. 45–70

Prescribing Institutions Without Ideal Theory*

David Wiens
Philosophy, University of Michigan

I. PRELIMINARIES

POLITICAL philosophers are not shy to prescribe ways to design social and political institutions so as to eradicate or at least mitigate various actual injustices. The conventional wisdom is that such prescriptions are the province of nonideal theory but that ideal theory is required as a guide for nonideal theory. On this view, our first task is to specify fully just *principles of regulation*—that is, principles that regulate the constitution of fully just institutional arrangements.[1] These principles subsequently guide and constrain our attempts to prescribe institutional solutions to address actual injustices.[2] Call this the *ideal guidance approach* to institutional design. This view is mistaken, or so I will argue. Ideal theory does not yield guiding principles that actual institutions must aim to realize, even if only approximately. Fortunately, the conventional wisdom is also avoidable.

Recent debate on the relationship of ideal theory to nonideal theory suffers from a dearth of clear alternatives to the ideal guidance approach. On the one hand, those who reject the notion that ideal theory is required as a guide for nonideal theory wind up rejecting ideal theory altogether as "normatively useless" without offering any alternative method for prescribing institutional reforms.[3] On the other hand, those who defend ideal theory against this charge wind up defending the notion that ideal theory *properly understood* is still useful as a guide for nonideal theory despite the limitations noted by critics, in part because there appears to be no other way to go about prescribing morally progressive institutional reforms.[4]

*Earlier drafts were presented to audiences in the philosophy and political science departments at the University of Michigan, as well as at the British Society for Ethical Theory Annual Meeting. Thanks to participants for the profitable discussions. I am particularly grateful to Elizabeth Anderson, Zev Berger, Steve Campbell, Bill Clark, Mika Lavaque-Manty, Peter Railton, and two anonymous reviewers for helpful comments. Support from the Social Sciences and Humanities Research Council of Canada (award no. 752-2007-0083) is gratefully acknowledged.

[1]My use of the term "principle of regulation" follows Cohen (2003, p. 241).

[2]Cf. Rawls 1999a, pp. 7–8, 215–16; 1999b, p. 89ff.

[3]See Geuss 2008; Farrelly 2007; Mills 2005. Amartya Sen (2009) is a partial exception here. He joins the others in rejecting ideal theory—or "transcendental theory", as he calls it—as being neither necessary nor sufficient for nonideal purposes, but goes beyond the others in offering a vaguely developed "comparative" methodological alternative.

[4]Robeyns 2008. Simmons 2010. Stemplowska 2008. Valentini 2009.

doi: 10.1111/j.1467-9760.2010.00387.x

To overcome this stalemate, we should pay less attention to characterizing the ideal/nonideal distinction and focus squarely on the following methodological question: *how should we approach the task of prescribing feasible institutional solutions to address actual injustice?* The issue here is identifying the relevant inputs as well as the appropriate procedure for turning those inputs into institutional design prescriptions. Although this issue is clearly related to debate concerning the ideal/nonideal theory distinction, I don't want to engage that debate directly. This is because "nonideal theory" is ambiguous between three different conceptions of the task of nonideal theory: (1) theorizing that identifies intermediate institutional reforms to help us transition from actual institutional arrangements to fully just institutional arrangements; (2) theorizing that identifies institutional arrangements that we should aspire to implement under actual conditions; and (3) theorizing that prescribes feasible institutional solutions to actual injustice.[5] Given the multifarious philosophical controversies surrounding that distinction, retaining the term "nonideal theory" allows this ambiguity to persist and potentially obscures my central question. Hence, to maintain focus on the central issue, I will refer to theorizing that prescribes feasible institutional solutions to actual injustice as *clinical institutional theory*, or *clinical theory* for short. To reiterate, the central issue here is how, methodologically speaking, we should approach the task of clinical theory.

Against the conventional widsom, I propose that clinical theorists should adopt an *institutional failure analysis approach*, which takes its primary design task to be obviating or averting social failures. The main innovation of this approach is to ground our evaluation of institutional arrangements on a detailed understanding of the causal processes that generate actual problematic outcomes. So conceived, failure analysis enables clinical theorists to prescribe more effective solutions to injustice because it focuses on understanding the *injustice*, rather than on specifying an ideal of justice. In so doing, failure analysis better fulfills the objective of clinical theory, namely, to think about how, in the midst of current injustice, we might bring about social conditions that are more just than our current conditions.

II. THE ARCHITECTURAL AND ENGINEERING PROBLEMS OF INSTITUTIONAL DESIGN

Institutions are, in Douglass North's familiar words, "the rules of the game in a society or, more formally, the humanly devised constraints that shape human

[5]We might call these (1) transitional theory, (2) nonideal aspirational theory, and (3) clinical theory respectively. Failing to disambiguate "nonideal theory" has had the effect of conflating the various conceptions. So, for example, it is not uncommon to find philosophers who think that doing clinical theory amounts to doing transitional theory. Indeed, this is what the ideal guidance approach suggests.

interaction".[6] More specifically, institutions are sets of (formal or informal) rules that establish roles and stabilize behavioral norms and expectations for occupants of those roles; these norms and expectations subsequently regularize patterns of interaction among individual agents. Since social outcomes arise from the aggregation of particular interactions, institutions are important for shaping social outcomes and their consequences for individuals' lives. For this reason, institutions are morally significant. Of course, an institution's effect on outcomes is not deterministic. An institutional structure sets the range of possible outcomes and makes some outcomes more probable than others. Thus, when assessing institutional arrangements, we not only care about the observed outcomes actually realized by institutions, but also the range and likelihood of outcomes they make possible.[7]

To the extent that we make choices over alternative sets of rules to regulate interactions, we design institutions. The task of clinical institutional design comprises two distinct sets of problems: architectural problems and engineering problems. To get some traction on this distinction, consider it in reference to a material structure such as a skyscraper or a bridge. The architect's objective is to design a structure that creatively organizes its components using mass, space, form, texture, and so on, in a way that embodies some set of values (often functional, economic, artistic, and aesthetic values).[8] The engineer's task is to apply mathematical and scientific principles to solve technical problems involved in the design of the structure within the constraints set by physical, technological, economic, environmental, and ethical considerations.[9] To be sure, the two tasks intersect—there is a discipline called "architectural engineering"—but, on the whole, architects understand their task as more closely aligned with art whereas engineers uniformly understand themselves as doing applied science.

Analogously, the architectural problems comprised by institutional design concern the ways in which different configurations of institutional components embody different sets of values. Here the salient values are likely to be functional, economic, and (of primary importance to moral philosophers) moral values. The engineering problems comprised by institutional design concern the application of social scientific principles with an eye to making institutions capable of withstanding the pressures to which they will be subject. Somewhat crudely, we can call the branch of institutional design that deals with architectural issues "applied ethics" and the branch that deals with the engineering issues "applied social science".

A key feature of the analogy is that the architectural and engineering problematics are mutually constraining as a result of their interdependence. A

[6]North 1990, p. 3.
[7]Thanks to an anonymous reviewer for bringing this point to my attention.
[8]Cf. Ching 2007, ch. 1; Roth 1993, ch. 1; Unwin 2003, ch. 1.
[9]Cf. Pahl et al. 2007, ch. 1; Holtzapple and Reece 2003, ch. 1.

generic example from the world of residential construction serves to illustrate the point. For any particular house we build, the architect gives us a set of drawings specifying the characteristics of the house—the location of the interior walls, the size of the window and door openings, the pitch and shape of the roof. For the most part, the architect confines herself to specifying the way the house is to *look*. When it comes to specifying the required structural features to make the architectural design work, the drawings are turned over to an engineer. Whereas the architect specifies the size of the openings, an engineer specifies the requirements for the "header" (a beam-like support placed above a window or door opening) to carry the load to be placed above the opening. The architect specifies the pitch and span of the roof, but it is a roof truss engineer who makes sure that the roof trusses are able to span the distance between load bearing points. If it is ever the case that a house can't be made to work structurally as the architect designed it—for example, the openings are too wide to accommodate adequate load bearing support—the drawings are sent back to the architect for modification. So the engineer specifies the structural limits within which the architect's design must work. But the engineer doesn't have a *carte blanche* when it comes to devising structural solutions; these are to be consistent with the architect's design aim as far as possible. If the architect calls for large, wide open living spaces, the engineer is not free to unilaterally add interior walls to accommodate a simpler roof truss design. Hence, the architect and engineer each set limits on the other's set of possible solutions for achieving their objective.

One way to characterize where conventional clinical theorizing has gone awry is to say that it has largely focused on applied ethics while paying insufficient attention to the relevant engineering issues. This is not to say that the applied ethics component is unimportant. Theorists have made some important progress in thinking about the moral values we want our institutions to embody and how different institutional configurations might embody different values. But if the analogy I've drawn is apt, then progress on the architectural problems is insufficient. To design feasible institutional solutions for unjust conditions, we must make progress on both the architectural and the engineering problems. This is because a structure that effectively embodies a chosen set of values but is incapable of withstanding the pressures to which it is subject ceases to embody the chosen values once it ceases to exist as designed.

Focusing almost entirely on the architectural issues has unfortunately recommended an institutional design procedure that limits the extent to which clinical theorists pay attention to the relevant engineering problems. Since philosophers have traditionally treated institutional design as an applied ethics problem, their design projects have typically focused on the application of ideal principles of justice to the design of actual institutions. This is why ideal theory has been thought to guide and constrain clinical theory. The basic intuition is that we want our actual institutions to bring about more just states of affairs. The

"more just" here intimates a target ideal according to which the justice of states of affairs can be measured. This suggests that clinical theory must seek ways to close the gap between current states of affairs and ideally just states of affairs. Hence, we require ideal theory to characterize a fully just institutional order, which serves as a *regulative ideal*, a guide to insure that our clinical theorizing is aimed in "the right direction". If a candidate proposal violates one or more of the principles of regulation identified by ideal theory or if it doesn't deliver institutions that more closely approximate the ideal, it is discarded. This ideal guidance approach leads us to judge design proposals according to their fit with the principles that regulate ideal institutional arrangements. Accordingly, we acquire a tendency to ignore the engineering question: how will the proposed institution fare in the face of the pressures to which it will be subject? The issue here isn't simply that clinical theorists are insufficiently attentive to important feasibility considerations. Feasibility considerations are but one set in the class of engineering considerations, which also includes stability and efficiency considerations. Moreover, even if we simply pay increased attention to feasibility, the ideal guidance approach leads our consideration of this issue to be circumscribed by the principles of regulation identified by ideal theory. We surely want our designs to be feasible, but our approach to design leads us to be primarily concerned with the extent to which our designs comport with ideal principles of regulation. This leads us to give short shrift to the engineering problems comprised by institutional design.

III. UNPACKING THE IDEAL GUIDANCE APPROACH: A CASE STUDY

Allen Buchanan's proposal to reform the international practice of recognition among sovereign states nicely demonstrates the ideal guidance approach in action.[10] Buchanan's proposal is an especially illuminating case study for two reasons. First, he clearly intends to prescribe a feasible solution to actual injustice.[11] Once we are assured that Buchanan *intends* his proposal to be so, we can ask whether he succeeds in prescribing reforms that are likely to "produce moral improvement in the particular system that now exists". Second, he self-consciously adopts the ideal guidance approach:

> The task of ideal theory is to set the most important and most distant moral targets for a better future, the ultimate standards for evaluating current international law. Nonideal theory's task is to guide our efforts to approach those ultimate targets. . . .[12]

[10]Buchanan 2004.

[11]"[W]e should eschew speculation about what constitutes a *comprehensive* set of ideal substantive institutional principles and concentrate on nonideal theory. . . . [W]e should focus on ascertaining which principles, if implemented, would produce moral improvements in the particular system that now exists" (ibid., p. 67).

[12]Ibid., p. 60ff.

The result is clinical theorizing that is assiduously constrained by ideal theoretic principles, which enables us to investigate the effect of this approach on Buchanan's nonideal prescription.[13]

The structure of Buchanan's book reflects his methodological orientation. Part one constitutes Buchanan's ideal theory. This comprises two basic theses concerning the design of the ideal international system.

(1) Institutions—in particular, the international legal system—must be designed to protect and promote *basic* human rights. These include a right to life, a right to physical security, a right against enslavement, a right to the means of subsistence, and a right against systematic racial, ethnic, or sexual discrimination.[14]
(2) International law ought to require nation-states to satisfy a minimal constitutional democracy condition.[15]

These theses support a justice-based conception of political legitimacy: a state exercises political power legitimately only if it respects basic human rights and is minimally democratic.[16]

In part two, Buchanan applies his ideal theoretic account of political legitimacy to reforming the institutionalized practice of sovereign recognition. The problem posed by the current practice is that it permits human rights-violating states to enjoy all the prerogatives of sovereign states, including "support for their territorial integrity and . . . noninterference in their internal affairs".[17] Hence, the current practice prevents us from adequately protecting individuals' human rights within the borders of rights-violating states. To preempt future rights-violators from using sovereignty as a shield against external interference, Buchanan proposes that international law be reformed to make a *new* polity's recognition as sovereign conditional upon satisfying four conditions:

(1) *Internal Justice Condition*: the state must protect (or must not violate) its citizens' basic human rights.

[13]Buchanan's characterization of nonideal theory here clearly resembles what I've called transitional theory. The quote in footnote 11 suggests that Buchanan also conceives of his proposal as part of clinical theory. In light of footnote 5, my point here implies that Buchanan has failed to distinguish between the different conceptions of nonideal theory and, thus, identified the task of clinical theory with the task of transitional theory. Since this is in effect what the ideal guidance approach recommends, this is further evidence that Buchanan has adopted that approach.

[14]Buchanan 2004, p. 129.

[15]Ibid., pp. 142–7. The arguments presented to support this second thesis derive it from the first thesis. Thus, (2) is, strictly speaking, a derivative thesis. But, as Buchanan notes, ideal theory comprises not only basic principles of justice that are to be satisfied by any institutional structure, but also "concrete principles that specify the institutional arrangements common to all systems . . . that satisfy the constraints laid down by the most basic principles of justice" (p. 67). Buchanan says that we are largely ignorant about principles of the second type with respect to the international legal system, with "one notable exception": that nation-states should be minimally constitutionally democratic. Hence, (1) is a basic ideal theoretic principle of the first type, while (2) is a basic ideal theoretic principle of the second type.

[16]Ibid., p. 234.

[17]Ibid., p. 266.

(2) *External Justice Condition*: the state must not violate the basic human rights of other states' citizens.
(3) *Nonusurpation Condition*: the state must not come about by usurping a legitimate state.
(4) *Minimal Democracy Condition*: the state must be minimally democratic.[18]

Each and every state that meets these conditions must be granted recognitional legitimacy; no state that fails to meet these conditions should be recognized.

To philosophers, Buchanan's argument looks just fine. He identifies a morally problematic feature of international law and suggests a solution that at least plausibly addresses the problem. So what's my objection to this approach? It's not that the ideal guidance approach yields the wrong verdict about the (in)justice of the current practice of recognition or identifies the wrong reasons for thinking that this practice is unjust. An alternative approach might ultimately arrive at the same verdict. Instead, the problem is that this approach at best arrives at an *incomplete* analysis, not least by emphasizing only a subset of the salient considerations for institutional design. Since international law issues from the activity and resolutions of states (as represented by the relevant government officials),[19] the feasibility and effectiveness of Buchanan's proposal hangs on its sensitivity to (at the very least) a host of considerations concerning both the likelihood and the depth of international cooperation on reforming the practice of sovereign recognition in a way that will drastically limit states' control over their internal affairs. Here I raise several considerations that Buchanan neglects and their implications for his proposal.

For Buchanan's proposal to have any effect on human rights performance, states must be willing to enact a binding resolution with enforcement provisions.[20] States are usually willing to bind themselves in this way only if the institutional mechanism is necessary to coordinate their activity to achieve a key policy objective.[21] One question, then, is whether states take human rights protection abroad as a foreign policy objective *that overrides competing policy objectives*. This seems implausible given even a cursory examination of the historical record. For example, US foreign policy history is checkered with support for rights-abusing dictators who were otherwise amenable to US foreign policy objectives, as well as operations to overthrow democratically elected

[18]On the first two, see ibid., pp. 269–72; on the third, see p. 275ff.; on the fourth, see p. 278ff.

[19]I typically use "state" as shorthand for "the government officials who are taken to represent a polity in international affairs". Consequently, the interests that matter are those of the state officials. Officials' decisions are influenced by their constituents' policy preferences *via* the domestic institutional mechanisms in place for holding state officials accountable. Where such mechanisms are robust, officials' policy decisions largely reflect the interests of the people at large. Where those mechanisms are weak, officials have more latitude in their policy decisions. Cf. Bueno de Mesquita et al. 2003; McGillivray and Smith 2008.

[20]See Hafner-Burton and Tsutsui 2007; Hathaway 2002, 2007.

[21]Cf. Keohane 1984; Milner 1997; McGillivray and Smith 2008.

governments who were deemed hostile to US interests.[22] In addition, rights-abusing states with great natural resource wealth continue to find support despite their human rights record, such as China's (among others') continued support for the Burmese military junta.

Rights-abusing states in particular will have little incentive to sign on to reforms that will effectively prevent them from recognizing new rights-violating states. To comply with the norm would only draw greater public attention to their own rights-violating practices and risk arousing domestic opposition that could be sufficient to drastically limit their ability to achieve their objectives. Given that some of the most internationally influential states are among the worst rights-abusers or most prominent supporters of rights-abusers (most notably, China and Russia), Buchanan's proposal is likely to meet stiff resistance among this crowd.

Moreover, many states that are relatively rights-respecting at home support rights-violating states abroad for a variety of reasons. Prominent among these reasons is the gains accrued from cooperation with such states. For example, many states with significant natural resource wealth turn out to be authoritarian, rights-abusing regimes.[23] Thus, otherwise rights-respecting states must cooperate with rights-abusing states to meet their substantial resource needs.[24] Given that similar benefits would accrue to states that cooperate with future rights-abusing regimes seeking recognition, states have little incentive to withhold recognition. Were they to do so, they would forego the substantial gains from cooperating with rights-abusing regimes.

Summing up: Buchanan's proposal neglects several considerations that are important for assessing the feasibility and effectiveness of his proposal. This is because he takes "justice . . . as the fundamental vantage point from which to evaluate the existing international legal system and to formulate proposals for improving it".[25] In other words, the set of considerations to which Buchanan gives adequate attention is restricted by the primacy he gives to his ideal principles of justice. He thereby fails to acknowledge where our interest in actually improving human rights protection might require us to make tradeoffs between justice on the one hand and feasibility and effectiveness on the other. To the extent that Buchanan's ultimate objective is to prescribe reforms that "would produce moral improvements" in our world, this is a serious blind spot. To

[22]Examples of the former include Fulgencio Batista, Mobutu Sese Seko, Pol Pot, and Saddam Hussein (before the late-1980s). Examples of the latter include Iran in 1953, Guatemala in 1954, the Democratic Republic of Congo in 1960, and Nicaragua during the 1980s.

[23]The literature on this so-called "resource curse" is abundant. See, for example, Ross 1999, 2001; Wantchekon 2002.

[24]Perhaps most important among these is the need for oil. Roughly 36% of the world's oil is produced by unquestionable human rights abusers (Saudi Arabia, Russia, Iran, and China), and of the other top-15 producers, Amnesty International reported serious human rights concerns in seven of them (US, Mexico, Venezuela, Kuwait, UAE, Nigeria, and Iraq). (Oil data from CIA 2008, human rights information from Amnesty International 2008.)

[25]Buchanan 2004, p. 73.

overcome this blind spot, our institutional design prescriptions must aim at more than simply approximating ideal principles of justice. We must prescribe solutions that are capable of overcoming social problems as we find them in the actual world.

IV. THE FAILURE ANALYSIS APPROACH TO INSTITUTIONAL DESIGN

We need an alternate approach to clinical theory that integrates applied ethics with applied social science. In this section, I develop an approach to clinical institutional design called *institutional failure analysis*, which takes averting failure to be the primary design aim. Failure analysis avoids the shortcomings of the ideal guidance approach by dispensing with the need for ideal principles of regulation to guide the design process. In the place of ideal theory, failure analysis emphasizes hypothesis formulation and evaluation. This places the design emphasis on overcoming actual social problems rather than on closing some gap between the actual and an ideal.

The motivating insight of failure analysis is well-stated by Henry Petroski:

> Desire, not necessity, is the mother of invention. New things and the ideas for things come from our dissatisfaction with what there is and from the want of a satisfactory thing for doing what we want done. More precisely, the development of new artifacts and new technologies follows from the failure of existing ones to perform as promised or as well as can be hoped for or imagined. Frustration and disappointment associated with the use of a tool or the performance of a system puts a challenge on the table: Improve the thing. Sometimes, as when a part breaks in two, the focal point for the improvement is obvious. Other times, such as when a complex system runs disappointingly slowly, the way to speed it up may be far from clear. In all cases, however, *the beginnings of a solution lay in isolating the cause of the failure and in focusing on how to avoid, obviate, remove, or circumvent it*.[26]

We see here a thumbnail sketch of failure analysis as a design process. The process starts with dissatisfaction, with a sense that some designed artifact doesn't work as well as we might like. It then proceeds to diagnose the problem: In what does the failure consist? What caused it? Upon analyzing the failure, the designer seeks to design something that will avoid the same fate by improving its capacity to withstand similar pressures, by removing exploitable weaknesses, or by constructing a design that is not subject to the same causal mechanism. Once the new design is complete, the designer tries to anticipate ways in which the new design might fail and, if any potential weaknesses are found, tries to improve the design to avoid these shortcomings.[27]

[26]Petroski 2006, p. 1 (emphasis added).

[27]Cf. Petroski 1992, p. 44. A problem arises here for the analogy, namely, that the intended "use" of institutional arrangements is usually much more contested than that of engineered artifacts. I address this difficulty below.

From this sketch, we can isolate three main phases of the failure analysis design process: (1) *identify* a failure (i.e., a flawed product or service); (2) *diagnose* the failure (i.e., analyze the character and cause of the failure); (3) *design* an artifact to overcome identifiable failures, including potential future failures. The design objective is, quite simply, to create things that avoid failure as far as is feasible. Developing the institutional failure analysis approach involves elaborating on each of these phases as they pertain to clinical institutional design. I now discuss these in turn.

A. Identifying Failure

An institutional design project is motivated by an initial sense of dissatisfaction with some feature of the social world, whether this is an observed undesirable outcome or social arrangements that impose an undue risk of realizing some undesirable outcome. However, an apparent disanalogy between engineering design and institutional design arises immediately. Failure analysis presupposes a well-defined design objective, which includes a set of general specifications that an artifact must meet. This well-defined objective makes it easy to identify failure. For example, a bridge should hold its intended load across the length of its span while withstanding environmental pressures, such as wind load or earthquakes. Bridges that fail to meet this design objective, or can only do so by incurring unacceptable costs, are readily identifiable as failures.

But social and political institutions do not come with well-specified design objectives. Indeed, politics is the process of contesting which ends are to be pursued by institutions. Consequently, social failures seem identifiable only by reference to some particular perspective. Asymmetric bargaining power in trade negotiations is a failure for those who inhabit a weak bargaining position, but the asymmetry rarely disturbs those who benefit from it. Income inequality is problematic for those at the bottom of the distribution, but rarely disturbs those who benefit from the institutions that lead to inequality. In general, this means that social failures are notoriously difficult to identify. The institutional failure analysis approach appears to suffer from an early setback.

This apparent difficulty gives life to the ideal guidance approach: how can we know when some feature of the social world counts as a failure unless we know how the social world *should* be constituted? It's natural to think that ideal theory is helpful here. Ideal theory yields a well-defined design objective; it identifies principles that serve as a general specification any institutional order must meet to count as just. With ideal principles of regulation in hand, failure identification becomes a simple matter: an institutional order whose principles of regulation fall short of or otherwise differ from the ideal counts as a failure. Since ideal theory derives principles of justice from a putatively *impartial* perspective, we need not worry about any particular perspective tainting our judgment. Ideal principles of justice give us the requisite impartial critical edge.

A clarification is in order before responding to this worry. "Ideal" ordinarily connotes something like "that to which we (ought to) aspire". Paradigmatic examples of ideals in this sense include equality, individual liberty, and human flourishing. With this in the background, my rejection of the ideal guidance approach is apt to be interpreted as the claim that clinical theory ought not appeal to ideals in this ordinary sense. Such an interpretation misunderstands my claim. Ideal theory is not "a theory of ideals", but a way of theorizing about political principles that focuses on specifying the principles of regulation that undergird an ideal institutional structure. These principles are arrived at by reflection on the principles that best express our moral ideals (in the ordinary sense) under ideal conditions, where "ideal conditions" denotes social and political circumstances that more or less permit moral considerations to take center stage in decisions about how to organize our collective life. For example, Rawls's difference principle is supposed to be the principle that best expresses our collective commitment to an ideal of society as a system of fair cooperation among citizens conceived as free and equal.[28] But the derivation of that principle makes idealizing assumptions to avoid complicating the moral analysis too much. These include, among other things, that society is self-sufficient and closed to transactions with outsiders, that citizens accept and know that others accept a common set of principles, and that citizens fully comply with the demands of the principles of justice.[29]

My rejection of the ideal guidance approach denies that the principles of regulation that express our commitment to our moral ideals under ideal conditions can or should offer any guidance for clinical theorizing. Importantly, this does not entail that clinical theorizing ought to refrain from appealing to moral ideals in the ordinary sense. In particular, it's open to the failure analyst to appeal to ordinary ideals, or *values* as I'll call them, when discussing the (in)justice of any particular social arrangements. To continue the above example, my rejection of the ideal guidance approach denies that Rawls's difference principle can or should offer any direction for our thinking about the justice of institutions that affect distribution under actual conditions. This is consistent with appealing to the underlying value of society as a system of fair cooperation among citizens conceived of as free and equal when assessing these institutions. This is because the difference principle is not a conceptual truth; we don't arrive at the difference principle by simply analyzing the concepts expressed by the value. The difference principle is a particular expression of that value, which follows from our reflection on that value *given a certain idealized conception of the political world*. To the extent that the derivation of that principle is sensitive to changes in initial conditions, its service as an expression of an important value

[28]Rawls 1999a, §§ 1, 3, 4.
[29]Ibid., pp. 4, 8, 216.

under different conditions will be in question.[30] But this problem affects only particular principles *qua* expressions of values, not the abstract values themselves. Thus, the latter remain available to the failure analyst when assessing institutional arrangements.

I now turn to a development of the failure analysis approach to identifying social failures. We identify failure by examining the particulars of the state of affairs that initially motivates the design project and then compare this state to alternative feasible states of affairs.[31] The contrast cases can be actual or counterfactual; the key is that their realization be feasible. If we take actual cases, we can be confident that the contrast class presents us with alternatives that are in some sense feasible, since they present states of affairs that are already realized. We need to be a little cautious when including counterfactual cases in the contrast class, since the fact that they are not currently realized leaves us uncertain about the extent to which their realization is feasible. However, this shouldn't preclude our making comparisons with counterfactual alternatives. We don't want our sense of which states of affairs are practically possible to be limited by what is actual. The general point is that we should be judicious in selecting our contrast cases, since these are going to determine which conditions to take as problematic and which to take as moral goals.[32]

A case about which there is reasonable disagreement will work best to illustrate my point here, so say we're dissatisfied with health care provision in the United States. For our set of contrast cases, we might select Canada, Cuba, Mexico, Nigeria, Russia, Sweden, and Tanzania. (I stick with actual cases for simplicity.) We start by making a rough intuitive ordering of these cases according to the relative justice of their health care provision schemes—for example (from best to worst): Canada, Sweden, United States, Cuba, Mexico, Russia, Tanzania, and Nigeria. This first cut ordering will be relative to some particular interpretation of justice, is likely to focus on some dimensions of health care provision at the expense of others, and is almost certain to be contested. This is fine for now; all we need is some set of orderings to serve as the raw material for the next step of failure identification.

[30]This point clearly echoes the basic idea expressed by Lipsey and Lancaster's (1957) general theory of second best: an institutional arrangement that is optimal under ideal conditions is unlikely to be so once we deviate from any of those conditions. The optimal arrangements under nonideal conditions are likely to require (perhaps drastic) alterations to the principles of regulation identified by ideal theory. Cf. Coram 1996; Goodin 1995.

[31]Cf. Sen's (2009) distinction between "comparative" and "transcendental"—i.e., ideal theoretic—approaches to justice. Although Sen identifies the crux of this distinction as being one of distinct objectives, the key difference between the two approaches is more accurately described as one of method. Transcendental theory employs the ideal guidance approach to identify principles to govern our selection of institutional arrangements. The comparative approach identifies these principles by examining and reflecting on our comparative judgments of actual or feasible social conditions. My development of the failure analysis approach explores this difference in method.

[32]More needs to be said about how to select counterfactual cases. Since feasibility is a workhorse concept here, its definition is crucial but beyond the scope of this paper. The relevant literature on feasibility includes the following: Brennan and Pettit 2005; Cowen 2007; Räikkä 1998.

The next step is to justify any particular ordering as capturing morally salient differences between the cases. At this point, one reflects on the considerations motivating any particular ordering and offers reasons for thinking that these considerations are (among) the morally salient ones when it comes to judging health care regimes. Example considerations include performance along objective health benchmarks (life expectancy, infant mortality rates, disease rates, etc.), health care spending efficiency, and scope of access to health care. Suppose one's ordering is primarily driven by a country's performance according to objective health benchmarks, as the above ranking is,[33] while another's ranking is primarily driven by access-related considerations. On behalf of the above ranking, one might say that health outcomes are key when judging health care regimes because a society should be primarily concerned with the objective well-being of its citizens and positive health outcomes are important markers of well-being. On behalf of access-related considerations, one might say that citizens, in virtue of their common status as citizens, are entitled to equal treatment in the allocation of health care resources. This is far from a complete characterization of what takes place at this point, but the picture being painted is sufficient to illustrate the point. Once we have a set of first cut orderings, we set about justifying an ordering as authoritative, which leads us to reflect upon the moral values we endorse and our reasons for endorsing them, as well as identifying the principles that best express those values. Moral justification requires us to engage in the process of supporting the moral authority of an ordering with impartial reasons—reasons that do not appeal to any particular person's situation or interests. Such a process will include appeals to abstract moral values. Importantly, this is not the same as identifying the principles of regulation that govern the ideal health care regime.

At some point, this debate will lead us to some shared judgments, although we are unlikely to arrive at complete consensus. For example, all parties to the debate might agree that health outcomes and access-related considerations are both important, although they might disagree on their reasons for thinking so or the relative weight assigned to each. No matter. This rough agreement still permits us to make judgments of the following sort: improving objective health outcomes in Nigeria and Tanzania constitutes an improvement in health care provision; increasing access to health care provision in the United States constitutes an improvement in health care provision. These verdicts imply judgments of failure. When we judge that some state of affairs *S* can be improved upon, we are committed to the claim that *S* is not as good as it could be. But not just that. Since our comparative judgments issue from our reflection on the moral values that underpin our judgments, our reasons for taking any ordering as morally authoritative imply that *S* is not as good as it *should* be. In Sen's words,

[33]The above ranking is according to life expectancy at birth, 2010 estimates. See CIA 2010.

these judgments identify "remediable injustices".[34] On the failure analysis approach, a failure just is a remediable injustice.

One might object that this makes failure an overly capacious concept, which undermines its critical edge. Perhaps *failure* should be reserved for social conditions we deem severe injustices requiring urgent attention. But failure need not be constituted by utter inability to meet design expectations. Instead, a failure is constituted by the presence of a remediable design flaw. We can comfortably acknowledge that instances of failure will differ along a number of dimensions, including ease of identification, severity, and (moral) urgency. A bridge collapse constitutes a greater failure than an unwieldy water bottle. Similarly, avoidable famine, genocide, total breakdown of the rule of law, and arbitrary detention and torture are more grievous and more urgent than, say, disparities in educational quality or employment opportunities. It's true that the generosity of the failure concept will preclude the identification procedure from generating fine-grained distinctions among failures, which could help us set remediation priorities. But that's not the job of *the identification procedure*. The identification procedure simply seeks to identify the members of the set of failures. Once we've identified (some of) the members of this set, we engage in further reflection and debate about our remediation priorities, debate that will and should appeal to considerations generated by our diagnosis of the failure and our anticipation of the effects of various intervention possibilities.

The preceding has exposed an important difference between the ideal guidance and failure analysis approaches. On the former, a social process or outcome is identified as a failure because it diverges from the processes or outcomes that would arise from a fully just institutional order. On the latter, there is no comparison with an ideal institutional order because there is no ex ante target institutional order. As a result, there is no preexisting blueprint from which actual institutions could diverge. Instead, we identify failures by making comparisons between actual and feasible states of affairs and finding that some actual states aren't as good as they could (and should) be. Consequently, we don't need a blueprint of the ideal institutional order to tell us which social conditions constitute failures.

In fact, such a blueprint is liable to bias our identification of failures because it prejudges what counts as a problem and thereby restricts our attention to certain features of the social world. Take Dewey's criticism of *laissez-faire* liberalism (i.e., libertarianism) as an example.[35] Libertarianism identifies individual liberty with individual economic enterprise more or less unconstrained by government regulation. The concomitant institutional ideal consists of an unregulated market for all goods in which people could have an interest. When a libertarian assesses the justice of an institutional order, her attention is restricted

[34]Sen 2009, p. vii.
[35]What follows draws from Dewey [1935] 2000, ch. 2, *passim*.

by her expectation that a just institutional order includes (at least) an unregulated market. The problem, in Dewey's words, is that libertarians have "put forward their ideas as immutable truths good at all times and places." Accordingly, the values of libertarianism have become reified: "[T]hey [hold] that beneficial social change can come about in but one way, the way of private economic enterprise, socially undirected, based upon and resulting in the sanctity of private property." This is due to a lack of "historical sense"; libertarians have been "blinded . . . to the fact that their own special interpretations of liberty, individuality and intelligence were themselves historically conditioned, and were relevant only to their own time". Once the sought-after reforms were accomplished, what was once a force for social change thus became a force in favor of the status quo. Consequently, adherence to the *laissez-faire* ideal blinds libertarians to the obstacles to effective liberty brought about by the market institution. Since their attention is directed by their favored institutional ideal, libertarians have failed to see the ways in which liberty is restricted by that very ideal.

The general lesson here is this: one doesn't need to know what the ideal institutional order looks like to be able to identify social failures. The ideal guidance view locates the blueprint at the wrong place, prior to the failure identification phase. But the creation of the blueprint is a *result* of the design process. There is no blueprint prior to the design phase, let alone the identification phase.

B. Diagnosing Failure

Once we've identified a social problem, we set about diagnosing that problem. This diagnostic phase incorporates both normative and empirical analyses. A return to the architecture/engineering analogy is useful here to get a grip on how the normative and empirical are intertwined. Architectural failures and engineering failures differ in virtue of the distinct design aims of architecture and engineering. Architectural design fails when space is poorly organized, or when component textures, materials, colors, and so on are poorly juxtaposed, or when the structure is a poor fit—functionally, artistically, aesthetically—with the surrounding environment. As examples, consider the structures that often show up on "ugly building" lists, such as Boston City Hall, the Experience Music Project (Seattle), and the Scottish Parliament Building (Edinburgh).[36] Engineering design fails when structural elements are unable to withstand environmental pressures, or when the object functions poorly or not at all under the conditions for which it was designed, or when the object is unsafe for use. Examples of engineering failures include bridge and building collapses, such as the Tacoma Narrows Bridge (1942) and the Hyatt Regency walkway (Kansas City, 1981). From the different lists of examples, it should be clear that architectural and

[36]See Steere 2008; Schiffman 2002; Virtual Tourist 2010.

engineering failures have distinct characters. Accordingly, the two require different sorts of diagnoses. When diagnosing architectural failures, we characterize the ways in which the object deviates from accepted standards or norms of architectural design. On occasion, architecture that was initially criticized is later seen as pushing the field in a positive, innovative direction. Hence, architectural analysis also involves re-evaluating the standards by which we assess architecture. In contrast, when diagnosing engineering failures, we identify the causal processes that generated the failure, in particular, the weaknesses in a structure or process and the pressures that were able to exploit that weakness.

Given that institutional design comprises both architectural and engineering problems, institutional failure analysis requires two types of diagnosis. Moral diagnosis identifies the features of an institutional structure that deviate from widely accepted norms and values, but also re-evaluates those norms and values. Causal diagnosis identifies the causal mechanisms generating the social conditions we seek to alter.

In practice, the tasks of moral and causal analysis are rarely separable. Our evaluation and selection of the moral principles by which we assess institutional arrangements will be informed by our causal analysis of current conditions. If income inequality is an unavoidable feature of collective economic life and we think collective economic life is important (or inevitable), then we might decrease the weight we give to particular egalitarian moral principles when morally assessing institutions. But if income inequality is a result of institutions that unnecessarily and unjustifiably restrict the economic opportunities of an underclass, we might retain those same egalitarian moral principles as important standards for institutional assessment.

Moral considerations also play an important role in identifying the causal mechanisms to which our causal diagnoses pay attention. For example, commodity price volatility almost certainly plays a causal role in generating the resource curse.[37] Thus, the independent market decisions of investors and consumers that are responsible in the aggregate for this price volatility are, at least in part, causally responsible for the misery associated with the resource curse. However, we don't typically identify these independent market decisions as causes of the curse and not simply because it's unlikely that we could adequately coordinate those decisions to avoid price volatility. In addition, we take it for granted that the market freedom that generates price volatility is a value that we should protect. Consequently, we turn our attention to other causal mechanisms. In the case at hand, we hold the fact of price volatility fixed and turn our attention to the mechanisms that make the economic performance of resource abundant states vulnerable to price shocks. All this is to say that moral and causal analyses are practically inseparable. Nevertheless, it can be useful to think of

[37]See Humphreys, Sachs, and Stiglitz 2007, ch. 1.

moral and causal analyses as analytically distinct. With this in mind, I now elaborate on my sketch of the diagnostic task.

Moral diagnosis characterizes the ways in which a social process or outcome constitutes a *moral* problem. This involves considering which values we want our social life to embody, why these values are important, and which moral principles best express those values. Do we prize equality? What about equality is important? Is individual liberty a key value? How is liberty restricted under current conditions? Although these questions inquire about abstract moral values and principles, on the failure analysis approach, we do not settle these questions solely in the abstract. Instead, our reflection on moral principles is guided by reflection on actual social conditions.

Consider Mill's vigorous criticism of Victorian marriage contracts as the sort of moral diagnosis I have in mind.[38] Mill begins his criticism in earnest by examining "the conditions which the laws of this and all other countries annex to the marriage contract".[39] Among other ills, such contracts left a woman effectively propertyless and thereby without economic security should her husband die or divorce her. Without any property of her own, she was financially dependent on her husband and without any credible exit threat should her husband abuse her. Women had no legal rights over their children; these were granted only to men. What's more, women were compelled into this position of servitude because they were banned from pursuing the means to independence, such as an education or a career. Mill compares the position of women to that of a slave:

> I am far from pretending that wives are in general no better treated than slaves; but no slave is a slave to the same lengths, and in so full a sense of the word, as a wife is. Hardly any slave, except one immediately attached to the master's person, is a slave at all hours and all minutes; in general he has, like a soldier, his fixed task, and when it is done, or when he is off duty, he disposes, within certain limits of his own time, and has a family life into which the master rarely intrudes. . . . But it cannot be so with the wife.[40]

The implicit argument here is that the Victorian marriage institution subjected women to a position that was deemed unfit for slaves and that the continued subjection of women to such conditions was inconsistent with earlier judgments against subjecting slaves to such conditions. In other places, Mill compares marriage to political tyranny, implying that the marriage contract subjected women to a position relative to their husbands to which no man would have consented in relation to a political ruler.

Mill's strategy involves enumerating the conditions that result from a particular institution and then exposing the conflict between these conditions and

[38]Mill [1869] 2002, esp. chs 2 and 3.
[39]Ibid., p. 153.
[40]Ibid., p. 155.

the moral principles we might justifiably endorse upon reflection. This results in both moral criticism of the conditions of marriage and a re-assessment of the convictions that keep those conditions in place. What's striking about Mill's strategy from the perspective of conventional political theory is what he doesn't say. Mill does not argue for a set of ideal principles of justice and then employ them to justify treating women as equals. Starting with abstract principles of justice permits us to rationalize concrete social conditions, to point to the ways in which the status quo is consistent with the requirements of these principles. This often requires minimizing (or altogether ignoring) inconvenient facts. By starting with concrete conditions, we cannot be let off the hook. We must come face-to-face with the details of our social reality and try to reconcile those details with our convictions. Often times, we cannot.[41]

Importantly, on the failure analysis view, the diagnostic phase involves not only identifying the ways in which current conditions undermine important values, but also *re-evaluating* the standards by which we assess social conditions. Moral principles are adopted in light of particular social conditions. Under conditions of inequality, particular egalitarian principles are endorsed; under conditions of slavery or tyranny, liberty is championed. But social conditions are in continual flux. Hence, "old principles [might] not fit contemporary life as it is lived, however well they may have expressed the vital interests of the times in which they arose".[42] The diagnostic phase demands that we reconsider our moral principles in light of our social reality to avoid adopting principles that are ill-suited to current conditions and to prevent principles of justice from becoming reified.

Keeping in mind the purely analytic distinction between moral and causal analysis, our causal analysis is an entirely empirical task. We are interested in explaining the outcome, not in assessing it according to normative criteria. Thus, *causal analysis* involves identifying the salient components of the causal process(es) that generate an outcome and specifying their interrelationships. This latter part includes specifying how the components interact and

[41]One might turn around and press this claim against my argument, namely, that failure analysis has a conservative bias. After all, on my view, coming face-to-face with social reality only leads to judgments of injustice (and thus a need to prescribe interventions) if we find features of that reality dissatisfying. Accordingly, we might follow G. A. Cohen (2003, p. 243) in claiming that "the question for political philosophy is not what we should do but what we should think, even when what we should think makes no practical difference". But Cohen's claim is consistent with my rejection of the ideal guidance approach and gives me a way to reply to this "conservative bias" charge. The failure analysis approach rejects the claim that ideal principles of justice are useful for guiding clinical institutional design. But it need not reject the claim that ideal theory can be useful for guiding *our attitudes* toward our social reality, including actual institutional arrangements. Finding gaps between actual institutional arrangements and ideal institutional arrangements motivates acute dissatisfaction with current arrangements, thereby initiating the design process. Thus, ideal theory can be important for *motivating* design prescriptions. This is consistent with the claim that the principles it yields ought not *guide* clinical institutional design. (Thanks to an anonymous reviewer for pressing me to address this issue.)

[42]Dewey [1927] 1954, p. 135.

how changes in one part of the process affect (the operation of) other components.[43]

Successfully incorporating causal diagnoses into our clinical theorizing requires philosophers to critically engage with the relevant social scientific research. Unfortunately, the quality of philosophers' interaction with social science is spotty and there seems to be no clear sense of how to most effectively incorporate social science into normative debate. This is not the place to give a full account of the methodological rules that should guide such interaction. Instead, I briefly outline some of the common missteps to help motivate clinical theorists' need to reflect more on how greater critical interaction with social scientists can enrich their institutional analyses, as well as why their institutional analyses require such interaction.

One key problem is that philosophers' typical mode of engagement with social science tends to treat the latter as providing ready-made, "off-the-shelf" answers to the scientific questions relevant to institutional design. This results in the widespread practice of simply citing social scientific results to support the empirical premises one wishes to employ. This type of passive reliance is problematic. First, it's bad practice. Not all studies are equal in quality. Many studies have poorly specified statistical models or unsuitable formal models, which bias the results of the study.[44] Citing biased results vitiates an argument for an institutional design proposal that presupposes the state of affairs portrayed by the biased results.

Second, a body of social scientific literature is not necessarily relevant to the specific institutional design question simply because it ostensibly addresses the same broad topic. There are two general pitfalls here, both of which discredit any argument that succumbs to them. One is that the data used in the cited empirical literature might be poor measures of the phenomena with which the normative literature is concerned.[45] The other pitfall is relying on an empirical literature that speaks to the effect of some intervention on one type of outcome when the normative debate is concerned with that intervention's effect on another type of outcome. For example, Mathias Risse argues that our duty to alleviate severe individual poverty generates a duty to assist poor countries' attempts to improve their domestic institutions on the basis of research that shows domestic

[43]Full description of what causal analysis involves requires, at a minimum, a social ontology, which identifies the fundamental objects of social causal processes; and, furthermore, an account of social causation, which says whether there are any social laws or mechanisms and specifies their respective roles in causal explanations of social phenomena. This is not the place to present such an account. For useful introductions to the relevant topics, see Elster 2007; Hedström and Swedberg 1998.

[44]Misspecification is more widespread than one would hope. For example, see Brambor, Clark, and Golder (2006) on the pervasive misspecification of interactive statistical models in the top political science journals.

[45]For example, Reddy and Pogge (2009) find extant measures of poverty lacking. But see Ravallion (2009) for a rebuttal.

institutional quality to be a key determinant of aggregate economic growth.[46] Aggregate economic growth might be related to individual prosperity, but not obviously so.[47] Whether economic growth leads to poverty reduction depends on how the gains from growth are distributed, and the gains aren't always distributed evenly.[48] The point is that Risse's argument would be much more compelling were he to appeal to studies with individual-level outcomes as their dependent variables. The content of our duty to poor individuals depends on what causes prosperity *for individuals*, not countries.

Third, political philosophers are too easily satisfied to rely on correlations identified by social scientists. The problem here is that an institutional design prescription proposes a way to intervene in and alter an existing causal process that is generating a problematic outcome. Thus, the salient question is: *what causal mechanisms generate the identified correlations?* In basing a prescription upon mere correlations, philosophers are in danger of prescribing, at best, interventions that alleviate symptoms instead of targeting the underlying cause, and, at worst, interventions that leave the operative causal process untouched. Thomas Pogge's discussion of the international resource privilege illustrates this point.[49] Pogge argues that the resource privilege is important for explaining the severe poverty found in resource abundant countries on the basis of two correlations: that between resources and low economic growth, and that between resources and authoritarian rule. To be sure, there are causal explanations of these correlations on offer.[50] But Pogge never concerns himself with the business of sketching, or even citing an account of the operative causal mechanisms. Such an account is what he needs to infer that the resource privilege harms the global poor. To conclude that the resource privilege harms the poor on the basis of these correlations, the causal logic that links resource abundance with civil conflict and authoritarianism must be of a piece with the causal logic that links resource abundance with reduced economic growth. What we need is an explanation that includes the causal processes linking resources, economic growth, and authoritarianism as part of a single, coherent, unified causal analysis.

The lesson here is twofold. First, causal analysis requires going beyond philosophers' typically passive modes of engagement with social science. Relatedly, successful clinical theorizing crucially involves diagnosing the problem we seek to overcome, a task that necessarily involves causal analysis. Thus, it's a weakness of the ideal guidance approach that it generates prescriptions that typically ignore the importance of adequate causal analysis.

[46]Risse 2005.

[47]See, e.g., Dollar and Kraay (2002) in support of the connection between aggregate growth and individual prosperity. For criticism, see Ravallion 2001.

[48]World Bank 2001.

[49]See Pogge 2002, pp. 113–4, 163–5.

[50]I discuss this issue in depth elsewhere (Wiens 2010).

C. Designing to Avoid Failure

Once we have a working diagnosis, we set ourselves to the design task. This balances design objectives that follow from both the moral and causal diagnoses. Normatively, the design aim is to prescribe institutional solutions that will bring about social conditions that comport with the moral values we can endorse on the basis of impartial reasons. The design aim that follows from our causal diagnosis is to prescribe feasible institutional solutions that can intervene effectively at important places in the causal process to improve the outcome.

Again, although we may analytically separate the normative and empirical design aims, in practice, they are tightly intertwined. Return again to the architecture/engineering analogy. Architects and engineers mutually constrain the design process. As my construction example in Section 2 illustrated, the physical limits of different structural materials and their various possible configurations constrain architectural possibilities; architectural standards and values constrain the set of desirable engineering solutions. Similarly, the moral principles we choose to endorse are constrained by the means required to realize those principles, while our assessment of interventions is constrained by the moral cost of implementing those interventions. To illustrate these points, consider the following (perhaps extreme) examples. If the realization of global income equality requires the establishment of a world government and the latter would entail unacceptable costs (moral or otherwise), the foregoing requires that we reconsider our endorsement of global income equality as an attractive moral aim. Similarly, if the best means to consolidating democratic governance requires drastically circumscribing the liberty of dissenters, the foregoing requires that we look for other ways to consolidate democratic reforms.

A key part of the design phase is design evaluation. We are not interested solely in overcoming *identified* failures; we also seek to forestall *identifiable* failures. Each institutional design proposal is a hypothesis that the institution as designed will successfully achieve its objective under the conditions in which it will be required to operate. Given the stakes, we can't accept such hypotheses blindly. Nor is intuition a reliable check. Thus, the last phase of the design process is to evaluate our design hypotheses for potential weaknesses and potentially negative path dependencies. Will the proposal generate morally perverse consequences? Will the institution be exploited by enterprising opportunists? Will it close off important possibilities for improvement in the future? Should we find weaknesses, we return to the drawing board to find ways to shore them up. If simple fixes aren't available, we need to consider overhauling the original proposal. The aim is to establish institutions that can foster and coordinate interactions in a way that, when aggregated, lead to morally improved social conditions that keep open possibilities for future improvement, as well as mitigate or contain the negative consequences of socially destructive interactions.

The importance of the evaluation phase and the willingness to revise in light of potential weaknesses suggests that clinical theorizing is a fluid, experimental process. We do not seek to propose an institutional configuration for all time. Nor do we aim to put together a "master plan" that encompasses an entire system of institutions. We are, of course, interested in uncovering the interactions between distinct components of a larger system and avoiding negative interactions as far as possible. But we should not hold out hope for a fully worked out ideal. Our vision is too limited, our knowledge too local. Each proposal is tentative and experimental, aiming at piecemeal, incremental progress.

None of the foregoing shows that the design process need not be guided by target states of affairs. Indeed, when stating the aim of the design phase, I claimed that we had "morally improved social conditions" in view. Doesn't this suggest the need for ideal theory to guide our thinking about what constitutes "morally improved social conditions"? To make the objection stronger, note that it need not rely on a view of ideal theory as delivering a singular best state of affairs. The easy reply in this case is that ideal theory is not up to the job.[51] All that's required to get the objection off the ground is that the design phase must be *prospective* in the sense that it requires a more or less well-specified target state of affairs and that ideal theory is required to identify the principles of regulation that govern the targeted arrangements. This is sufficient to vindicate at least a restricted version of the ideal guidance approach.

The fact that clinical theory seeks to bring about a moral improvement of social conditions suggests that clinical institutional design must aim at *something*, namely, morally improved social conditions. But on the failure analysis approach, the design phase is largely *retrospective* in the sense that our sights are set by looking backward, at the places we've been rather than at the places we'd like to go to. We design institutions to avert failure, not to realize an ideal. To be sure, the evaluation phase requires that we try to anticipate potential future failures and we might say that this makes failure analytic design prospective. But this is not prospective in the sense used by the ideal guidance approach. More accurately, it is *counterfactually* retrospective. That is, the evaluation phase examines where we would have been had we implemented the original design proposal. Our design aim then is to overcome these counterfactual failures.

Designing to overcome failures has no need for the principles of regulation identified by ideal theory. All we need to know is (1) which possible solutions are feasible; (2) which of the feasible solutions are morally acceptable; and (3) which of the feasible, morally acceptable solutions are likely to effectively intervene at the appropriate place in the causal process generating the failure. None of this makes reference to the principles of regulation identified by ideal

[51]See Sen 2009, ch. 4.

theory. One might argue that we need ideal theory to identify the morally acceptable solutions within the feasible set. But there's no reason to employ the overwrought framework of ideal theory to help us here. Attempting to identify ideally just arrangements and the principles that regulate them is liable to distract from the task of identifying solutions. In any case, ordinary moral reasoning is sufficient for identifying which feasible options are morally acceptable. Return to the earlier discussion of the comparative method for identifying failures (Section 4.1). A similar method is effective here too. We take our current social conditions and compare them to the conditions that would arise were we to implement some particular feasible institutional solution. We then ask ourselves: Do we think the counterfactual conditions are acceptable? Are the counterfactual conditions an improvement upon current conditions? On the basis of which principles do we make these judgments? Can we justifiably endorse these principles upon reflection? This comparative process needs nothing like ideal theory to identify principles of regulation to serve as targets for clinical theory. The resources we need to prescribe morally progressive institutional solutions are available without having to take on board the baggage of ideal theory.

V. CONCLUDING REMARKS

I've argued that we should abandon ideal theory in our attempts to prescribe institutional reforms to address actual injustices and adopt a failure analytic approach to clinical institutional design. But the preceding development of the failure analysis approach has remained largely abstract. It remains to be seen how substantive clinical theorizing about particular problems can be improved by adopting this approach. The best way to do this is to employ the failure analysis framework in thinking about problems like extreme poverty and inequality, war and military intervention, and gender and racial discrimination. For obvious reasons, I must leave this for another time.[52] Nonetheless, the foregoing is sufficient to outline the key differences between failure analysis and the conventional approach to clinical theorizing. The primary drawback of the ideal guidance approach is its myopia; it focuses on understanding and applying an ideal of justice at the expense of a detailed understanding of the problem. Failure analysis overcomes this myopia by refocusing our attention on the problem. By integrating empirical (causal) analysis with moral analysis, a failure analytic approach is more sensitive to the complexities of the problems we wish to address without sacrificing sensitivity to the important moral considerations that rightly constrain our attempts to address our social failures.

[52]But see Wiens (2010) for an application of the failure analysis approach to clinical theorizing about the resource curse.

REFERENCES

Amnesty International. 2008. *Report 08: The State of the World's Human Rights*. London: Amnesty International. Available at: <http://thereport.amnesty.org/eng/Homepage> (last accessed July 24, 2008).

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. Understanding interaction models: improving empirical analyses. *Political Analysis*, 14, 63–82.

Brennan, Geoffrey and Philip Pettit. 2005. The feasibility issue. Pp. 258–79 in Frank Jackson and Michael Smith (eds), *Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press.

Buchanan, Allen. 2004. *Justice, Legitimacy, and Self-Determination*. New York: Oxford University Press.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson, and James D. Morrow. 2003. *The Logic of Political Survival*. Cambridge, MA: MIT Press.

Ching, Francis D. K. 2007. *Architecture: Form, Space, and Order*, 3rd edn. Hoboken, NJ: Wiley.

CIA. 2008. Rank order—oil production. *The CIA World Factbook*. Available at: <www.cia.gov/library/publications/the-world-factbook/rankorder/2173rank.html> (last accessed July 25, 2008).

CIA. 2010. Rank order—life expectancy at birth. *The CIA World Factbook*. Available at: <www.cia.gov/library/publications/the-world-factbook/rankorder/2102rank.html> (last accessed October 16, 2010).

Cohen, G. A. 2003. Facts and principles. *Philosophy & Public Affairs*, 31, 211–45.

Coram, Bruce Talbot. 1996. Second best theories and the implications for institutional design. Pp. 90–102 in Robert E. Goodin (ed.), *The Theory of Institutional Design*. New York: Cambridge University Press.

Cowen, Tyler. 2007. The importance of defining the feasible set. *Economics and Philosophy*, 23, 1–14.

Dewey, John. [1927] 1954. *The Public and Its Problems*. Athens, OH: Ohio University Press.

Dewey, John. [1935] 2000. *Liberalism and Social Action*. Amherst, NY: Prometheus Books.

Dollar, David and Aart Kraay. 2002. Growth is good for the poor. *Journal of Economic Growth*, 7, 195–225.

Elster, Jon. 2007. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. New York: Cambridge University Press.

Farrelly, Colin. 2007. Justice in ideal theory: a refutation. *Political Studies*, 55, 844–64.

Geuss, Raymond. 2008. *Philosophy and Real Politics*. Princeton, NJ: Princeton University Press.

Goodin, Robert E. 1995. Political ideals and political practice. *British Journal of Political Science*, 25, 37–56.

Hafner-Burton, Emilie and Kiyoteru Tsutsui. 2007. Justice lost! The failure of international human rights law to matter where needed most. *Journal of Peace Research*, 44, 407–25.

Hathaway, Oona A. 2002. Do human rights treaties make a difference? *Yale Law Journal*, 111, 1935–2042.

Hathaway, Oona A. 2007. Why do countries commit to human rights treaties? *Journal of Conflict Resolution*, 51, 588–621.

Hedström, Peter and Richard Swedberg, eds. 1998. *Social Mechanisms*. New York: Cambridge University Press.

Holtzapple, Mark T. and W. Dan Reece. 2003. *Foundations of Engineering*, 2nd edn. New York: McGraw-Hill.

Humphreys, Macartan, Jeffrey D. Sachs, and Joseph E. Stiglitz, eds. 2007. *Escaping the Resource Curse*. New York: Columbia University Press.

Keohane, Robert O. 1984. *After Hegemony*. Princeton, NJ: Princeton University Press.

Lipsey, R. G. and Kelvin Lancaster. 1957. The general theory of second best. *The Review of Economic Studies*, 24, 11–32.
McGillivray, Fiona and Alastair Smith. 2008. *Punishing the Prince: A Theory of Interstate Relations, Political Institutions, and Leader Change*. Princeton, NJ: Princeton University Press.
Mill, John Stuart. [1869] 2002. *The Subjection of Women*. Pp. 123–230 in Dale E. Miller (ed.), *The Basic Writings of John Stuart Mill*. New York: Random House.
Mills, Charles W. 2005. "Ideal theory" as ideology. *Hypatia*, 20, 165–84.
Milner, Helen V. 1997. *Interests, Institutions, and Information*. Princeton, NJ: Princeton University Press.
North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
Pahl, Gerhard, Wolfgang Beitz, Jorg Feldhusen, and Karl-Heinrich Grote. 2007. *Engineering Design: A Systematic Approach*, ed. and trans. Ken Wallace and Lucienne Blessing, 3rd edn. London: Springer.
Petroski, Henry. 1992. *To Engineer is Human: The Role of Failure in Successful Design*. New York: Vintage Books.
Petroski, Henry. 2006. *Success Through Failure: The Paradox of Design*. Princeton, NJ: Princeton University Press.
Pogge, Thomas W. 2002. *World Poverty and Human Rights*. Malden, MA: Polity Press.
Räikkä, Juha. 1998. The feasibility condition in political theory. *Journal of Political Philosophy*, 6, 27–40.
Ravallion, Martin. 2001. Growth, inequality, and poverty: looking beyond averages. *World Development*, 29, 1803–15.
Ravallion, Martin. 2009. How not to count the poor? A reply to Reddy and Pogge. Pp. 86–102 in Sudhir Anand, Paul Segal, and Joseph Stiglitz (eds), *Debates on the Measurement of Global Poverty*. Oxford: Oxford University Press.
Rawls, John. 1999a. *A Theory of Justice*, rev. edn. Cambridge, MA: Harvard University Press.
Rawls, John. 1999b. *The Law of Peoples*. Cambridge, MA: Harvard University Press.
Reddy, Sanjay G. and Thomas W. Pogge. 2009. How *not* to count the poor. Pp. 42–86 in Sudhir Anand, Paul Segal, and Joseph Stiglitz (eds), *Debates on the Measurement of Global Poverty*. Oxford: Oxford University Press.
Risse, Mathias. 2005. What we owe to the global poor. *Journal of Ethics*, 9, 81–117.
Robeyns, Ingrid. 2008. Ideal theory in theory and practice. *Social Theory and Practice*, 34, 341–62.
Ross, Michael L. 1999. The political economy of the resource curse. *World Politics*, 51, 297–322.
Ross, Michael L. 2001. Does oil hinder democracy? *World Politics*, 53, 325–61.
Roth, Leland M. 1993. *Understanding Architecture*. Boulder: Westview Press.
Schiffman, Betsy. 2002. The world's ugliest buildings. *Forbes*, May 3, 2002, <www.forbes.com/2002/05/03/0503home.html> (last accessed October 16, 2010).
Sen, Amartya. 2009. *The Idea of Justice*. Cambridge, MA: Harvard University Press.
Simmons, A. John. 2010. Ideal and nonideal theory. *Philosophy & Public Affairs*, 38, 5–36.
Steere, Mike. 2008. The world's ugliest buildings. *CNN International*, November 7, 2008, <http://edition.cnn.com/2008/WORLD/europe/10/22/ugliest.buildings/> (last accessed October 16, 2010).
Stemplowska, Zofia. 2008. What's ideal about ideal theory? *Social Theory and Practice*, 34, 319–40.
Unwin, Simon. 2003. *Analysing Architecture*, 2nd edn. New York: Routledge.
Valentini, Laura. 2009. On the apparent paradox of ideal theory. *Journal of Political Philosophy*, 17, 332–55.
Virtual Tourist. 2010. The world's top 10 ugliest buildings. Available at: <http://members.virtualtourist.com/vt/t/1c7/> (last accessed October 16, 2010).

Wantchekon, Leonard. 2002. Why do resource abundant countries have authoritarian governments? *Journal of African Finance and Economic Development*, 5, 57–77.
Wiens, David. 2010. Natural resources, political development, and the demands of justice. Unpublished manuscript, University of Michigan, <www.davidwiens.org>.
World Bank. 2001. *World Development Report 2000/2001: Attacking Poverty*. New York: Oxford University Press.