On the Nature and Possibility of Ideal Guidance

by

Charles de la Cruz

A Senior Honors Thesis
Submitted to the Department of Political Science
University of California, San Diego

April 2, 2012

Contents

Key Terms and Abbreviations

NT (Theory) **Nonlinear Transitional Theory**
A method of deriving guidance from institutional ideals by evaluating policy alternatives based not on immediate similarity to the ideal, but on the preservation of the ability to fully realize the ideal in the future. This is the central concept of this essay.

Linear Transitional Theory
An approach to ideal guidance in which immediate alternatives are ranked based on similarity to an institutional ideal. It is frequently and erroneously considered to be synonymous with the concept of ideal guidance, and as a result the crippling flaws of linear transitionalism are often held to be problems with ideal guidance in general.

(Strictly) Comparative Theory
A method of promoting the development of just institutions through the selection of policies based on preference rankings of various immediately achievable alternatives. By "strictly comparative" I mean that it foregoes attempts to derive guidance from ideal theorizing. This is the main alternative to ideal guidance.

PD Tree **Path Dependence Tree**
My method of visualizing branching possibilities over time. Essentially a specialized decision tree that aids in illustrating the mechanisms of path dependence and transitional modes of planning. A PD tree lays sideways and uses both the vertical and horizontal axes. The x-axis denotes chronological progression (left to right), and the y-axis denotes increasing short-term desirability (bottom to top).

ToJ ***A Theory of Justice*** **by John Rawls**
A foundational text in contemporary political theory. I draw frequently on Rawls's analysis of various aspects of justice and ideal theory throughout the essay.

LoP ***The Law of Peoples*** **by John Rawls**

TMS ***The Theory of Moral Sentiments*** **by Adam Smith**

List of Figures

Acknowledgements

My sincerest thanks to Gerry Mackie for being a fantastic thesis advisor, striking a perfect balance between guiding my thought and trusting me to chase after my own intuitions—even when it wasn't clear where they would lead.

I am also forever grateful to Fonna Forman, my teacher and mentor. It is under her guidance these past years that my study of political theory has grown from an interest to a passion, and from a passion to a calling.

Chapter I
Ideal and Nonideal Theory

> *…this law and its predecessors are all fine. But I think, Socrates, that if we let you go on speaking about this subject, you'll never remember the one you set aside in order to say all this, namely, whether it's possible for this constitution to come into being and in what way it could be brought about.*
>
> Glaucon, *Republic*[1]

The tension between what we would today refer to as ideal and nonideal theory has existed since the beginning of political philosophy itself. On the one hand we, as humans, are capable, through reason and reflection, of imagining and designing ideal societal arrangements that embody conceptions of perfect justice, whatever they may be. On the other, in our actual political affairs we are perpetually condemned to fall short of the requirements of these ideal societies due to constraints both internal and external, social and psychological, fixed and malleable. The ever-present question we are left with when confronted with this disconnect is the same one Glaucon put to Socrates over two thousand years ago: if an ideal theory of a perfectly just society is far beyond the reach of our present capabilities, *what good is it?*

[1] Plato, *Republic*, trans. G.M.A. Grube (Indianapolis: Hackett Publishing Company, 1992) 471c

Throughout the history of normative political philosophy, those who would theorize about the possibilities of political arrangements have had to take into account this inescapable conflict between what we can imagine and what we can presently do. Discussion of the issue can be found across the ages, from Plato to Machiavelli[2] to Rousseau[3] to—most significantly for the current state of debate—the late John Rawls. It is Rawls who formulated the contrast between ideal and nonideal theory in the terms used in contemporary discussions. In the pages that follow, I will take as a starting point the Rawlsian framework of constraints and assumptions that separate ideal from nonideal theory. However, departing from most contemporary debates on the value or utility of ideal theory, I mount a defense of ideal guidance based not on its questionable and primarily intuitive necessity, but on its practical methodological value.

For the first time in history we are witnessing the emergence of non-imperial global institutions[4] that can provide a tangible foundation for international claims of justice. But we are at the same time facing serious shortcomings and growing imbalances in wealth, quality of life, and political and economic power. The question of whether or not the current state of inter- and intra-national affairs is characterized by deep injustices is not particularly contentious. The issue at stake in discussions of distributive justice is not whether the world is unjust, but rather what

[2] "For many have imagined republics and principalities that have never been seen or known to exist. However, how men live is so different from how they should live that a ruler who does not do what is generally done, but persists in doing what ought to be done, will undermine his power rather than maintain it." Machiavelli, *The Prince* (Cambridge: Cambridge University Press, 2008) p. 54

[3] In the first paragraph of *The Social Contract*, Rousseau takes note of nonideal limitations in his stated plan of "taking men as they are and laws as they might be."

[4] Or, at the very least, not *necessarily* imperial.

can be done about it. If ideal conceptions of justice, no matter how fair or argumentatively sound, are to have any practical value (and I argue they should), they will have to be taken beyond the pages of academic journals to become a part of a broader process of public deliberation and understanding. But if ideal theory is to be taken seriously by people and policy, its value will have be formulated in practical and nonideal terms. In advocating a 'practical' account of ideal theory's usefulness I do not mean to say that ideal theory must move toward the empirical quantification that has characterized the development of social science over the last century—I mean merely to say that it is not enough to develop an ideal theory of justice that *should* guide nonideal decisions. We who would advocate the pursuit of a distant ideal must also show *how* it should guide present choices. *The demonstration of a specific method and mode of thought through which ideal theory can inform and improve nonideal decisions about what policies and institutional arrangements to pursue here and now is the task of this essay.* A complete account of ideal guidance would be a massive undertaking indeed, amounting essentially to an attempt to bridge the ancient gap between political philosophy and political action. While the limited nature of this essay precludes such a complete account, I hope to lay the foundation for a new way of thinking upon which such an account could be built.

The essay proceeds in four parts. In this first chapter, the concept of ideal theory is carefully analyzed and placed in its historical and contemporary intellectual contexts. I also explore competing ways of thinking about the progress of justice in the world in order to clarify where the present argument fits within recent

debates. The second chapter deals with the epistemological problems created by the need for "second-best solutions." In a nonideal world that seriously constrains the accessibility of ideal arrangements, evaluating which of the immediately available options we should consider to be second best is a task even more difficult than one might initially assume. It is in response to the difficulties of second-best solutions that I develop the idea of *nonlinear transitional theory* as a method of evaluating alternatives with reference to an institutional ideal while avoiding, to some extent, the crippling uncertainty associated with such choices. Chapter 3 builds on this foundation and develops a fuller account of how this method of evaluation based on ideal guidance can be realized. It does so through a parallel discussion of the concepts of *path dependence*, in which future alternatives are limited by present choices, and *feasibility*, which explores what it means for an ideal to be "possible" as well as how one might go about progressing toward a feasible ideal. The fourth and final chapter brings the main line of argument to a close before branching out to touch on various other aspects of ideal theorizing that must also be considered if a complete account of ideal guidance is to be made.

In §1 of this chapter**,** as an introduction for the unfamiliar, I provide an outline of John Rawls's distinction between ideal and nonideal theory within his theory of justice. This will prove useful in providing a basic context for the discussion of justice and ideal theory to follow. §2 focuses on comparing and analyzing contemporary arguments about the role and value of ideal theory. Although the question at the root of this issue is ancient, sustained critical attention to it in a post-

Rawls world has emerged only in the last several years. As a result of this novelty, significant confusion and disagreement remains about both the implications and the definitions of the concepts employed in discussions of ideal theory. An organization of the present literature and terminology is necessary for further productive development of these ideas. Finally, §3 provides a brief summary of the essential ideas to be drawn from the chapter, as well as an outline of how the argument for the existence of productive applications of ideal theory in an imperfect world will be structured in the chapters to follow.

1. Rawls's Theory of Justice

John Rawls's *A Theory of Justice*, published in 1971, is a landmark work in political philosophy. With its publication came a lasting resurgence in the popularity and visibility of normative political theory, and it has profoundly shaped the landscape of debate about (among other things) justice and the actions it might require for the last 40 years. Justice, like time[5], is an idea that most people feel familiar with in an abstract sense until they are asked to pin it down with words. In early childhood many children are already making judgments about what is "fair," even as some of the greatest minds in history struggle and fail to capture the

[5] "What then is time?" asks Augustine of Hippo. "If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know."

essential character of justice. It is easy to take advantage of the general acceptance of the importance and intuitive necessity of justice and simply charge into the subject assuming the reader will follow close behind, armed with a similar definition of the term. However, in resisting the temptation to assume a shared groundwork I hope not only to clarify the terms on which the discussion is based, but also to contribute to the arguments that follow through an examination of the elements that constitute modern liberal ideas of social justice. Importantly, I do not attempt to defend the content of Rawls's theory of justice. Drawing from the more general understanding he develops of the basic nature of justice, and particularly on the division of a theory of justice into ideal and nonideal theory, I attempt instead to defend the value of ideal theory as a tool for realizing *any* particular set of feasible ideals.

As a final preliminary point, the prevalence of Rawlsian thought in contemporary political theory means that even when his ideas and their derivatives are not explicitly engaged, they form an implicit shared background knowledge. One can certainly follow recent discussions on justice and ideal theory without having read any Rawls, but placing these arguments in a broader intellectual context is just as important in the present as it is when reading more distant historical authors. I hope to provide in this section at least a rough sketch of such a context.

1.1: The Nature of Social Justice

"Justice," Rawls states on the first page of *A Theory of Justice* (hereafter *ToJ*), "is the first virtue of social institutions, as truth is of systems of thought."[6] In this brief but powerful statement lies the essence of a major trend among modern understandings of the nature of justice. In my interpretation, this analogy to truth is particularly enlightening for the issue at hand as it highlights what I consider two fundamental characteristics of Rawlsian justice: it is, like truth, both *binary* and *systematic*.

By binary I mean that in light of a particular conception of justice, an institution or set of institutions must be categorized as one of two possible options: just or unjust. This is not to imply that there is no *moral* distinction between various states of injustice; rather, such distinctions cannot be made with reference to a scalar measure of *justice*. Justice is not a quantity; it is a specific institutional arrangement along with the set of principles that define such an arrangement. Importantly, however, arrangements that diverge from the requirements of a particular conception of justice cannot be meaningfully ranked as more or less just based on their *similarity* to the ideal arrangement for reasons we will later explore. This idea that there is no linear scale for ranking unjust arrangements (which we inevitably face for the foreseeable future) forms a major part of recent critiques of ideal theory and will be central to the argument presented in chapter 2. The binary

[6] John Rawls, *A Theory of Justice* (Cambridge, Massachusetts: Harvard University Press, 1971) (hereafter *ToJ*) p. 3

categorization draws an important distinction between justice and other metrics of evaluation such as height or utility. Both height and utility can be seen as purely scalar metrics in which comparisons can be made between different magnitudes without reference to any final state of complete height, or maximum utility. These metrics have no endpoint that can be referenced when comparing alternatives; two measures of height can and need only be compared to each other. There is no absolute "high" that either measurement could be compared to.

Truth and justice, on the other hand, are binary (that is, categorical) in that they both have a "complete" state that does not allow for scalar comparison. Such limitations on scalar measurement can be understood in two parts: ranking just states and ranking unjust states. By treating justice as a particular institutional arrangement, rather than as something that a state of affairs has some quantity of, the idea of ranking two *just* states becomes meaningless. Similarly, two true statements cannot be ranked in regards to their truth. They are both, simply, *true* in an absolute sense. Any deviation that could differentiate them in this regard would have to make one of them *not* true. So it is with justice. As for ranking unjust states, treating justice as an arrangement rather than a quantity requires that any comparison of two unjust states be made with reference to the ideal of justice. This would, however, require some method of creating a linear ranking of deviation from just arrangements which, as we will see, creates serious problems.

The second quality highlighted by the analogy between truth and justice is that they are *systematic*; the appropriate state of each individual element depends

on all of the others. Whether or not a single institution is just depends not on the institution itself, but on how it fits into the complete arrangement. An institution that appears just in isolation may have a different form as a part of a complete just arrangement. As we will see, this has profound implications for the applicability of partial conceptions of justice and piecemeal solutions to injustice. For any given social arrangement, evaluations of whether or not it is just must take into account the interrelatedness of each part of the complete system. This idea of interrelated ideal conditions will also play an essential role in the development of the second chapter.

Another idea that can be unpacked from Rawls's opening statement is an explicit delineation of the realm of justice. The idea that justice is a virtue specifically of *institutions* has profound implications for the role of justice in the world. The inseparability of justice and institutions forms an important part of Rawls's specifically political conception of justice in which any claim to justice requires shared political institutions and, consequently, a common sovereign power. This reading is borne out by the fact that while Rawls's theory of justice deals with the basic structure of institutional arrangements within a state under a common sovereign power, his exploration of the possibilities of an international order tellingly does not treat the idea of justice as applicable to actors (nation states) lacking a common sovereign power.[7] Although some might view this Hobbesian requirement of sovereignty for social justice as aiming too low in the pursuit of

[7] John Rawls, *The Law of Peoples* (Cambridge, Massachusetts: Harvard University Press, 1999) (Hereafter *LoP*)

justice on a global scale, it is an inescapable limitation of the necessity of an institutional framework upon which claims of justice can be made.[8] In other words, the obligations of justice are political rather than moral; they emerge not from our shared humanity but from our shared social institutions.

I have so far been treating a conception of justice as an institutional ideal, yet someone familiar with Rawls might recall that his conception of justice famously takes the form of two *principles* of justice.[9] While it is true that the development of Rawlsian justice in the hypothetical 'original position'[10] begins with a decision regarding principles by which justice can be judged, he goes on to note that "having chosen a conception of justice, we can suppose that they are to choose a constitution and a legislature to enact laws, and so on, all in accordance with the principles of justice initially agreed upon."[11] The decisions made in the original position, i.e. ideal theory, include decisions not only on principles of justice but also on the institutions such principles would judge to be perfectly just. Thus while ideal theory includes broad evaluative principles, it also necessarily develops these principles into an agreeable institutional structure. Throughout this essay I, along with many of its

[8] For an insightful account of the global limitations of a political conception of justice see: Thomas Nagel, "The Problem of Global Justice," *Philosophy & Public Affairs* 33 (2005): 113-147

[9] The first principles states that "each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all." The second that "social and economic inequalities are to be arranged so that are both (a) to the greatest benefit of the least advantaged" and "(b) attached to offices and positions open to all under conditions of fair equality of opportunity." *ToJ* p. 302

[10] Rawls's justification for the validity of the two principles is that they would be unanimously preferred to all presently known alternatives by participants in a hypothetical "original position." While the exact nature of the original position emerges through the reflective equilibrium of the participants, Rawls presents a case in which every member of society has an equal say and judges from behind a "veil of ignorance" where they have no knowledge of their actual attributes, preferences, or position in society. *ToJ* pp. 136-137

[11] *Ibid.* p. 13

detractors[12], treat the output of ideal theory as an institutional arrangement rather than a set of principles for evaluating justice.

However, it must be noted that an institutional ideal of justice was, for Rawls, far from a complete social ideal. Rather than capturing in its entirety the irreducibly broad and complex set of social institutions that define a society, the institutions developed in Rawlsian ideal theory merely define what he calls the "basic structure" of society. This basic structure is defined by Rawls as "the way in which major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation" and is made up of the "political constitution and the principal economic and social arrangements."[13] The idea of justice itself obviously extends beyond the basic structure of society, and as a result a theory of justice for major social institutions "may be irrelevant for the various informal conventions and customs of everyday life."[14] But because an understanding of justice in every aspect of human life is a prohibitively large and complex task, Rawls (and most others) accept that "the basic structure is the primary subject of justice because its effects are so profound and present from the start."[15]

It is this basic structure that contains various social positions into which people are born, some naturally more desirable than others. These different positions, each leading to different life expectations, "affect men's initial chances in

[12] It is this understanding that prompts Amartya Sen to refer to Rawlsian ideal theory disparagingly as "transcendental institutionalism," as we will later see.
[13] *ToJ* p. 7; "legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family are examples of major social institutions."
[14] *Ibid.* p. 8
[15] *Ibid. p. 7*

life; yet they cannot possibly be justified by an appeal to the notions of merit or desert."[16] This basic structure of society, then, is taken as the focus of contemporary theorizing about justice (including this essay) because (1) it defines, perhaps more than any other social factor, the life expectations of individuals in society and (2) it is deeply affected by public policy choices, and can thus be shaped through public deliberation and political action.

1.2: A Two-Part Theory of Justice

With this contextual groundwork laid, the nature of ideal theory and its nonideal counterpart can be understood much more clearly. Rawls makes a distinction early on in the *ToJ* between what he at first calls strict and partial compliance theories[17], and later refers to as ideal and nonideal theory. His stated purpose in drawing a distinction between ideal and nonideal theory is to "split the theory of justice into two parts" in which "the first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances. It develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the fixed constraints of human life."[18] The essential idea here is that ideal theory is built upon an assumption of strict individual and institutional compliance to the ideals derived in this stage. Importantly, this assumption is both (a) not true in any real world

[16] *Ibid.*
[17] *Ibid.* p. 8
[18] *Ibid.* p. 245

society and will probably remain that way for the foreseeable future and (b) inseparable from the concept of ideal theory. Putting aside the specific mechanisms through which Rawlsian ideal theory is derived, in this context it is in the nature of ideal theory (of any kind) to be derived from nonfactual assumptions about a society's present capabilities. Such assumptions are precisely what make a theory ideal in the first place; a goal that was immediately achievable would be a policy proposal rather than an ideal. Such immediately achievable goals, constrained by present conditions, are instead the subject matter of *non*ideal theory.

Before moving on to nonideal theory, however, there is another essential aspect of Rawlsian ideal theory (and, as I will argue, utilizable ideal theory in general) that is frequently overlooked: it is *not* utopianism in the normal sense of an abstract and unachievable ideal that requires perfectly rational or benevolent human beings, the elimination of resource scarcity, or other such impossibilities. Ideal theory certainly abstracts from present limitations—especially on the subject of compliance—but it remains, and must remain, what Rawls calls a *realistic utopia* which takes (in the words of Rousseau) "men as they are and laws as they might be."[19] "Political philosophy," he says, "is realistically utopian when it extends what are ordinarily thought to be the limits of practicable political possibility and, in doing so, reconciles us to our political and social condition."[20] The nature of a realistic utopia will be explored more completely in chapter 3. For now, however, it is important to keep in mind that the argument for the possibility of ideal guidance

[19] *LoP* p. 7
[20] *Ibid.* p. 11

in the chapters that follow takes as its starting point an ideal that may not be achievable now, or even soon, but is still *realistically possible.*

But although his work deals almost exclusively with ideal theory, Rawls by no means believes this ideal alone to be sufficient for decision-making in the real world: "nonideal theory, the second part, is worked out after an ideal conception of justice has been chosen; only then do the parties ask which principles to adopt under less happy conditions."[21] In the nonideal stage of a theory of justice, the question shifts from "what does a just society look like?" to the much more immediate "what are we to do to improve our visibly unjust society?"

"Obviously," writes Rawls, "the problems of partial compliance [i.e. nonideal] theory are the pressing and urgent matters. These are the things that we are faced with in everyday life."[22] Any reasonable evaluation of policy alternatives must accept the immediate factual limitations that ideal theory puts aside. The day-to-day policy judgments we must make and the lives we lead are inescapably not ideal, and ideal theorizing cannot replace such fully contextual judgments. What remains, then, is the question presented at the beginning of this essay: what good are the institutional plans of ideal theory if our actual nonideal conditions prevent us from implementing them? And further: why is separating a theory of justice into two parts even necessary at all?

Although Rawls focuses primarily and deliberately on the content of ideal theory rather than the methods of applying it, he does address briefly the purpose

[21] *ToJ* p. 245
[22] *Ibid.* p. 9

and applicability of ideal theory in nonideal decision-making. "The reason for beginning with ideal theory," he says, "is that it provides, I believe, the only basis for the systematic grasp of these more pressing [nonideal] problems…at least, I shall assume that a deeper understanding can be gained in no other way."[23] Ideal theory, then, may not be able to prescribe specific policy proposals which necessarily depend on nonideal contexts, but it can at the very least provide us with a deeper understanding of the nature of the injustices we face and the possibility of overcoming them. This classically Rawlsian claim, that institutional ideals can be used to develop certain "systematic" understandings of nonideal choices which cannot come from anywhere else, is essentially the basis for the thesis of this essay. Yet despite the fact that an account of exactly how such systematic understanding could emerge would obviously have to be made at some point if ideal theory were to be used in practice, no substantial effort has been made to build that bridge. But where Rawls and others have been satisfied with the claim that ideal theory *should* be applied in some way, I aim at demonstrating *how*.

Rawls appears to make the above claim based on intuition rather than with any developed methodology in mind. After addressing this issue in passing near the very beginning of the *ToJ,* he returns much later to the unresolved lacuna between ideal and nonideal theory to present two brief and unsatisfying accounts of the applicability of ideal theory. The apparently weaker of the two is the claim that "if we have a reasonably clear picture of what is just, our considered convictions of

[23] *Ibid.*

justice may fall more closely into line even though we cannot formulate precisely how this greater convergence comes about. Thus while the principles of justice belong to the theory of an ideal state of affairs, they are generally relevant."[24] This account, appealing entirely to intuitions about the apparent dialectical value of ideal theory in bringing about the convergence of individual convictions, is far too vague and argumentatively groundless to be taken very seriously, at least in the way it is presented. It does, however, have some potential merit with further refinement, which I will touch on in the final chapter.

The second account of ideal guidance in nonideal circumstances takes the form of an appeal to the possibility of evaluating arrangements based on their relative *distance from the ideal*. This approach is significant in that it is frequently a major part of criticisms against the possibility of usefully applying ideal theory, and as such it is worth quoting Rawls fully:

> Viewing the theory of justice as a whole, the ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason…thus, as far as circumstances permit, we have a natural duty to remove any injustices, beginning with the most grievous as identified by the extent of deviation from perfect justice.[25]

[24] *Ibid.* p. 246
[25] *Ibid.*

This understanding of the method through which ideal theory can provide insights into real world decisions is the basis of what I call *linear transitional theory*.[26] First, it is transitional in that it takes as its goal the actual realization of the ideal. That is, it takes the institutional ideal to be realistically achievable and attempts to provide insight into how to achieve it. Second, it is linear in that it is based on the idea that institutional arrangements can be evaluated according to some absolute measure of "the extent of deviation from perfect justice." The immediate implication of such an understanding is that any given arrangement can be ranked either cardinally or ordinally against others along a single dimension. One could imagine this as a horizontal line with the ideal at one extreme, on which any particular institutional arrangements could be placed and compared to others. The idea, then, would simply be to choose the alternative that is closer to the ideal in any situation and hope that eventually a progression of such choices would bring a society in line with the ideal. The problems with such a scalar measure of justice have been mentioned briefly already, and will be developed further in §2.3 of this chapter.

Linear transitional theory, as noted above, is frequently taken to be the singular method through which ideal theory could be applied and as a result is the target of much (justified) criticism. But, in his defense, Rawls was not strongly

[26] The idea of *transitional* justice used here is adapted from A. John Simmons's "Ideal and Nonideal Theory," *Philosophy & Public Affairs* 38 (2010): 5-36. The term is applied, as he notes, "rather differently than its more common use to refer to the ways in which states address past human rights violations (during their transitions to social stability)."

committed to such an approach. After his above statement he quickly adds that "of course, this idea is very rough" with the further caveat that "the measure of departures from the ideal is left importantly to intuition."[27] Furthermore, there appears to be evidence that he envisioned a method of using ideal theory that did not depend on strictly linear progress, though it was never developed very far. In the *ToJ* he states at one point that nonideal restrictions on the equality of liberty are only acceptable "to the extent that they are necessary to prepare the way for a free society…so that in due course these freedoms can be enjoyed."[28] This is, consistent with his earlier statements, clearly a transitional way of thinking. But in this allowance for the restriction of liberty for the sake of of the long-term realization of the full liberties of ideal theory lies the seed of a different kind of transitionalism—a transitionalism in which the measure of progress is not immediate social benefit or greater similarity to the ideal, but instead the preservation of the ability to achieve the ideal in the future—even at the short term expense of greater benefits or similarity. The alternative way of approaching the prospect of ideal guidance hinted at here will later be developed much more extensively into the central concept of this essay: *nonlinear transitional theory.*

[27] *ToJ* p. 246
[28] *Ibid.* p. 152

2. Debating Ideal Theory

Recently, the role of ideal theorizing in the development of justice in society has come under increased criticism. With regard to this criticism, the issue of justice in political philosophy appears somewhat unique among philosophical subjects with respect to the demands made of it. As Adam Swift has observed, there is an apparent distinction between an "epistemological" and a "practical" study of justice.[29] The epistemological pursuit of truths about the concept of justice, independent of their applicability to the world, is criticized on the grounds that it provides no practical guidance for *promoting*, rather than merely *understanding* justice.[30] The underlying normative implication of such criticism appears to be that a theory of justice that, either directly or indirectly, guides action in the real world is *better* than a theory that does not. Such a theory is, of course, undeniably more useful for informing particular actions. But such practicality doesn't necessarily make it more *true*. There is certainly a tradition, which includes modern thinkers such as G.A. Cohen and stretches as far back as Socrates[31], of seeking out truths about justice without attempting to draw normative prescriptions for action, or even desiring that it be

[29] Adam Swift, "The Value of Philosophy in Nonideal Circumstances," *Social Theory and Practice* 34 (2008): 363-387

[30] "It is striking," Swift notes, "that we are less likely to criticize violinists, say, than political philosophers, for failing to provide justice-promoting guidance, as if being interested in identifying truths about justice meant that one was more rather than less culpable for failing to tell us how to bring it about." *Ibid.* p 367

[31] "Do you think that someone is a worse painter if, having painted a model of what the finest and most beautiful human being would be like and having rendered every detail of this picture adequately, he could not prove that such a man could come into being?...do you think that our discussion will be any less reasonable if we can't prove that it's possible to found a city that's the same as the one in our theory?" Plato, *Republic* 472d

possible to do so. It is hardly fair, then, to criticize these theories for failing to do what they consciously choose not to attempt.

My purpose in stressing this point is to clarify that what I am attempting to do in what follows is not to defend the value of ideal theorizing as an end in itself. Instead, I want to defend a particular species of ideal theory that, while operating under assumptions that do not represent current or immediately achievable conditions, can still be fruitfully employed in guiding action. I make a case not for ideal theory as a path to knowledge alone, but for ideal theory as a path to *practical* knowledge. Such a defense of the value of imagining a distant but *realistic* utopia must be based, as the content of subsequent chapters will be, on a demonstration of the method through which such practical knowledge and guidance could be derived. I begin, in this preliminary chapter, by working toward a clear and thorough definition of ideal theory. In the subsequent subsections I survey recent attempts to justify and to dismiss ideal theory in light of the need for a way of deriving practical guidance. Each subsection focuses on a particular dichotomy that illuminates various points of contention in contemporary theorizing about justice.

2.1: The Input-Output Distinction

Zofia Stemplowska, in a recent essay, provides a simple and useful framework for understanding current debates on ideal theory.[32] In her analysis, the

[32] Zofia Stemplowska, "What's Ideal About Ideal Theory?" *Social Theory and Practice* 34 (2008): 319-340

nature of ideal theory is understood within a broader framework of normative theory in general. The purpose of a theory, in the most general sense, is to take a set of inputs and, through a system of rules, derive a set of outputs. Few would take issue, I expect, with this broad formulation. In a positive theory data is input and, through some procession of rules, translated into predictions of what is or will be the case. In a normative theory, at least one value-based claim is input and a normative output is derived. In addition to general outputs about what should be the case (such as institutional ideals), normative theories may also provide outputs that identify specific policies or actions as desirable— "recommendations" as she calls them. Note that normative theorizing in this sense encompasses both ideal and nonideal theory. The systematic nature of a normative theory means that criticism of the value of ideal theory may be directed at either the input, the rules, or the output. Despite the slightly cold, almost mechanical nature of this characterization of normative theory as a simple three-step process of input-rules-output, I expect that keeping this fundamental framework in mind during the following survey of the landscape of contemporary discussions of ideal theory will prove beneficial.

The assumption of perfect compliance in Rawlsian ideal theory is clearly not empirically true (or likely to be true in the foreseeable future); it falls under what Onora O'Neill describes as *idealization.*[33] Yet although this assumption is the classic example of what makes a theory ideal, it need not be the only one. With a more

[33] Within normative theory, O'Neill classifies *abstraction* as theorizing based on conditions that are generalized or simplified, but not false. *Idealization* goes beyond abstraction and makes judgments based on untrue assumptions. From: *Towards Justice and Virtue* (Cambridge: Cambridge University Press, 1996)

general categorization of inputs based on this idea of idealization, an ideal theory might be understood to be any normative theory that takes as inputs assumptions that are presently and demonstrably false. However, Stemplowska notes, the mere existence of nonfactual inputs is not sufficient for a sound critique of the applicability of ideal theory; the fact that a theory's input includes false assumptions alone does not automatically invalidate the practical value of the output. A theory might take perfect compliance as an assumption but come up with an ideal (i.e. an output) that is directly achievable here and now. Conversely, a theory may transform a set of strictly factual inputs into an output with no practical value whatsoever.

The issue of the limitations on valid inputs for a normative theory of justice is far from settled[34], but I will put it aside for the current examination because, as Stemplowska succinctly states, "focusing on inputs when drawing boundaries between different types of theory is important only to the extent that we can show that putting something into a theory, or not putting something into it, will have an effect on the function of the theory."[35] That is, when the usefulness of ideal theory is in question, the most important distinction to make between different types of theory is how they can actually be utilized (i.e. their outputs). Although arguments may well be made that an idealized input necessarily leads to an output without value, any attempt at a resolution of such an argument must ultimately be made by examining the output and judging whether or not it really is useless.

---

[34] See: Colin Farrelly, "Justice in Ideal Theory: A Refutation," *Political Studies* 55 (2007): 844-864; Charles W. Mills, "'Ideal Theory' as Ideology," *Hypatia* 20 (2005): 165-184

[35] Stemplowska (2008) p. 324

Given this emphasis on outputs, Stemplowska attempts to pinpoint the pivot around which the debate over ideal theory turns by offering a new definition of ideal theory. Turning toward the outputs of normative theory she proposes that "a straightforward way of drawing the ideal/nonideal distinction is to define nonideal theory as a theory that issues AD[achievable and desirable]-recommendations, and ideal theory as theory that does not."[36] Such an understanding embraces the possibility that the recommendations of ideal theory cannot be directly applied to the nonideal world, but seeks to defend its value anyway. In doing so, she sketches the second of the two central arguments in the theoretical defense of ideal theory. These are (1) that an idealized assumption (input) does not necessarily undermine the usefulness of the conclusions (output) and (2) that even if ideal theory (using her modified definition) does not issue recommendations directly applicable to the real world, it still has value.

In my attempt to untangle the knot of ideas present in the recent literature on this subject, the input-output distinction serves as a useful tool in understanding and evaluating the role of ideal theory. I have arranged the possibilities presented so far into a grid (on the following page) to aid in visualizing the relationship between the different dimensions in which a theory may be classified as ideal.

[36] *Ibid.*

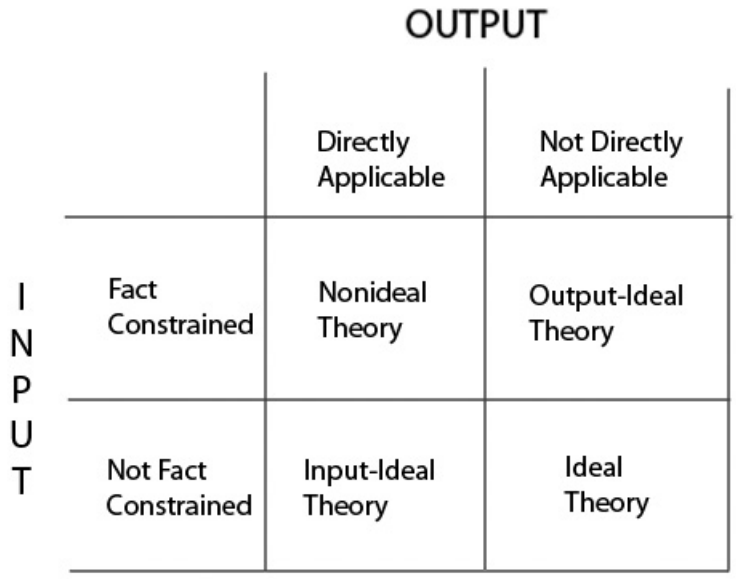

Fig. 1

Directly applicable here refers to whether or not the output of a given theory is able to provide direct (that is, unmediated by further theorizing) guidance when applied to a given nonideal state of affairs, and that this guidance is able to prescribe achievable and desirable courses of action. Such prescriptions would recommend policies or arrangements that are immediately achievable with our present abilities and constraints.

Judging whether or not an input is *fact constrained* creates significant epistemological questions regarding our limited knowledge of human psychology and sociology, as well as the difficulty of judging precisely which nonideal constraints must be respected and which can be treated as malleable. A theory that was fact constrained in an absolute sense would be forced to accept the status quo,

treating all present conditions as insurmountable factual limitations. Obviously, a normative theory must be able to treat *some* facts as non-limiting if any progress is to occur. These problems are closely related to the issue of *feasibility* in assessing which present limitations are *absolute* limitations. This topic will be examined thoroughly in chapter 3. For the time being, it is sufficient to observe that while inputs of fact constrained theories are thought to be constrained by all necessary factual limitations (but certainly may miss some or treat some conditions as constraints that actually are not), the inputs of theories that are not fact constrained contain things known, or at least generally accepted, to be presently false or beyond our abilities. When a theory's input and output are evaluated along these dimensions, two provisional definitions of ideality may be stated as follows:

(1) Ideal theory is a normative theory in which the input includes at least one untrue assumption. The outputs may be directly applicable or they may not. (The bottom row of boxes in the above diagram)
(2) Ideal theory is a normative theory that creates outputs that are not directly applicable. The inputs may be fact constrained or not. (The right column of boxes)

The first definition is derived from Rawls, the second from Stemplowska. However, while Rawls clearly defines ideal theory based on a nonfactual input (using "ideal theory" and "strict compliance theory" interchangeably), he does not assume it to be directly applicable to nonideal circumstances in all (or perhaps even

any) cases. The extent to which nonideal theory is a necessary second step is left ambiguous. Yet in stating clearly that the development of a just society requires both ideal and nonideal theorizing, Rawls endorses the necessity of nonideal theory in at least some, if not all, circumstances. Thus, while Rawlsian input-ideal theory may in fact be directly applicable in some circumstances, it is certainly not so in every case; he is clearly not placing himself in the bottom left box of strict input-ideal theory (fig. 1 p. 24), in which the results of ideal theory can be directly applied to any problem. While an infinitely insightful, one-size-fits-all ideal theory that is directly applicable to every situation without a need for further nonideal theorizing may exist somewhere, no one I have read or encountered is attempting to argue for it. This essentially eliminates pure input-ideal theory as an option.

Looking at the upper right corner (of fig. 1), a theory that was constrained by nonideal circumstances but still failed to create directly applicable outputs could be described as simply a failed attempt at nonideal theorizing. It would retain the limitations of nonideal theory (fact constraints), without the benefits (directly applicable outputs). A further application of a more complete theory would be necessary for specific prescriptions to emerge, making a pure output-ideal theory superfluous. With both pure input-ideal and pure output-ideal theories off the table, a complete definition of ideal theory, drawn from the bottom right box, may be stated as follows:

Ideal theory is a normative theory in which the input includes at least one nonfactual assumption and the output is not directly applicable to present nonideal circumstances.

With this in mind, criticism of the value of ideal theory can be understood most clearly as a debate about the process by which normative theorizing proceeds from basic value judgments to specific actions or policy decisions. For a proponent of ideal theory as defined here, the progression goes as follows:

(i) Input 1: basic values, assumptions (including untrue assumptions)

(ii) Application of rules of theory 1 ("ideal theory")

(iii) Output 1: Conception of ideal justice (not directly applicable)

(iv) Input 2: Factual constraints, conception of justice (from output 1)

(v) Application of rules of theory 2 ("nonideal theory")

(vi) Output 2: Specific policy or course of action to pursue (directly applicable)

A critic of the use of ideal theory would generally prefer the following progression:

(i) Input: basic values, justifiable assumptions, factual constraints

(ii) Application of rules of theory ("nonideal theory")

(iii) Output: Specific policy or course of action to pursue

As the list of steps here illustrates, the function of ideal theory is to translate basic value judgments into a conception of justice that can then be used as an input for a nonideal theory. The debate turns, then, on whether or not the additional step that ideal theory creates between basic moral values and specific policies can be justified. Recent critics of ideal theory argue that basic value judgments are sufficient inputs for a nonideal comparative evaluation without mediation, making ideal theory, at best, redundant.

2.2: Ideal Theory and Normative Ideals

Adam Swift, in a recent article, attempts to defend the value of abstract theorizing by pointing out the substantive difference between the output of a given ideal theory (that is, some conception of a perfectly just society) and the inputs and processes that produce it. He describes these latter parts as consisting of the *reasons* behind the principles that ideal theory designs. "As long as philosophers can tell us *why* the ideal would be ideal, and not simply *that* it is, much of what they actually do when they do 'ideal theory' is likely to help with the evaluation of options within the feasible set."[37] It is certainly true that the formation of the outputs of an ideal theory necessarily takes into account value judgments from which it derives a conception of a perfectly just society. This is uncontroversial. But by stepping away from the institutional outputs of ideal theory and defending instead the usefulness of

[37] Swift (2008) p. 365

carefully examined values, Swift has in actuality made an argument *against* the necessity of ideal theory as it is understood here.

The argument presented falls victim to a blurring of the line between two concepts that must be distinguished in a clear account of the function of ideal theory: general normative ideals and ideal theory. Ideals, on the one hand, are clearly necessary for any normative theory, be it ideal or nonideal. Reflecting on and carefully examining these foundational values is unquestionably important. However, if it is only the careful consideration of values and not the specific outputs of ideal theory that are of use, then there is no reason why a nonideal theory could not simply take those values into account along with factual constraints to create directly applicable outputs without first constructing a systematic account of perfect justice. This line of argument has been taken up by Amartya Sen and David Wiens, and will be examined in the following subsection. Thus, while Swift certainly makes a strong argument for the importance of philosophical reflection in understanding the values we use to evaluate our actions and institutions, his argument does not demonstrate the value of ideal theory itself.[38] Any value claim is capable of serving as a basis for nonideal judgments without appealing to an ideal theory. Thus, any claim that ideal theory is *necessary* for guiding policy will not be supportable.

[38] To be fair, the title of his article refers to the value of *philosophy* rather than of ideal theory. But attempts to defend ideal theory by appealing to the value of philosophy within the essay are ineffective.

2.3: Comparative and Transcendental Justice

Perhaps the most well known critic of ideal theory is Amartya Sen, who in recent years has argued that ideal theory (particularly Rawlsian ideal theory) is, as he puts it, "neither necessary nor sufficient" for relieving injustice in the world.[39] Sen's argument, as the most prominent strand of recent criticism of ideal theory, warrants close examination. In place of the conceptions of a perfectly just society developed in ideal theory, which he refers to as *transcendental institutionalism,* Sen proposes an alternative *comparative* approach to nonideal justice. This comparative approach does not attempt to identify perfectly just societal arrangements and then use those arrangements to evaluate present conditions and potential actions, as ideal theory might. Instead, it focuses only on comparative judgments of immediately achievable arrangements as "more" or "less" just than others without considering what perfect justice might look like.

It should be noted that in the two-stage process of ideal and nonideal theorizing that Rawls and others propose, comparative judgments are necessarily made among feasible options in the nonideal stage. Action of any kind requires *some* sort of comparative evaluation of some alternatives as "better" than others. These judgments would, however, be informed by an institutional ideal. Thus, the significance of Sen's argument comes not from a claim that comparative judgments should serve as a basis for practical decision making (few would dispute their

[39] Amartya Sen, "What Do We Want from a Theory of Justice?" *The Journal of Philosophy* 103 (2006): 215-238

importance), but from the claim that they can be made without any appeal to a transcendental (i.e. ideal) standard.

The first of Sen's claims, that ideal theory is not a *sufficient* standard for comparative evaluation, challenges the idea that an institutional ideal is capable of providing an unambiguous and linear scale along which any particular arrangement can be placed and ranked. If such a scale existed, then simply evaluating distance from the ideal would be sufficient for picking the "most just" out of a set of policy options. Against this notion, Sen rightly notes that any society that falls short of ideal arrangements will do so along multiple dimensions. Imagine, he might say, attempting to rank, in terms of distance from the ideal arrangement, a society with free speech but no right to privacy against a society with limited free speech and strong privacy protections. The comparison is difficult, if not impossible, and is certainly not unambiguous or objective. It is made even more difficult by the need to rank them not according to personal preference, which is at least feasible, but in accordance with some objective scale based on the ideal arrangement in which both rights are present. The plural nature of social institutions thus makes a complete linear ranking of social arrangements impossible with the result that a single ideal can never be sufficient to directly inform decision making.

Sen goes on to argue that ideal theory is, beyond being insufficient, not even a *necessary* part of comparative judgments. If establishing a consistent metric of distance from the ideal were possible, then Sen's sufficiency argument would not hold. However, given the impossibility of such a metric, it is unclear how an ideal of

justice would be of any use whatsoever in ranking two alternative arrangements. Sen illustrates this point through an analogy to height. The knowledge that Everest is the tallest mountain in the world, he says, "is neither needed, nor particularly helpful, in comparing the heights of, say, Kanchenjunga and Mont Blanc."[40] That is, when ranking two feasible alternatives, one can evaluate which is better without appealing to what is best (but not an option). Sen's arguments, particularly as illustrated in his analogies, raise important questions relating to certain fundamental issues regarding the concept of justice itself. For example: if justice is not defined by an ideal, what does it mean to say that any one arrangement is more "just" than another? Is justice then merely the immediate preference of the individual or group making any particular comparative judgment? These questions will be taken up shortly in the following subsection. But for now, if these questions can be put aside in favor a broader, albeit more vague idea of what justice entails, Sen's argument appears sound.

David Wiens, in line with this comparative approach, has presented a more developed account of the process by which comparative judgments might be utilized to design just institutions without appealing to institutional ideals. Filling in the gap left by Sen as to the actual process of comparative decision-making, Wiens offers an approach that he calls *institutional failure analysis*.[41] Prefacing his approach with a clarification of present usage, he presents a useful account of the various possible understandings of nonideal theory:

---

[40] Sen (2006) p. 222

[41] David Wiens, "Prescribing Institutions Without Ideal Theory," *The Journal of Political Philosophy* 20 (2011): 45-70

> "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.[42]

The first definition describes the role of nonideal theory in the Rawlsian two-part framework: ideal theory first identifies perfectly just institutions and then nonideal theory takes into account factual constraints to guide existing institutions toward the ideal. This is the essence of transitional theory. The second definition can be understood either with or without ideal theory. It may, on the one hand, take an idea of justice into account when deciding what the more limited "best" institutional alternative should be under present constraints. This is, however, certainly not necessary. This second type of nonideal theory may simply attempt to design a set of institutions to aspire to based on general normative principles and immediate limitations. The third definition appears to be what Sen and Wiens have in mind in their comparative approaches. For them, the function of nonideal theory is not related to any specific institutional goal. It focuses instead on the identi-

[42] *Ibid.*

fication and alleviation of immediate and concrete injustices. While it does still appeal to "ordinary moral reasoning," its primary task is, as Wiens puts it, "obviating or averting social failures."[43]

Extending the architectural metaphors so frequently employed in discussions of ideal and nonideal theory, I would like now to take a moment to sketch an analogy that hopefully clarifies the basic intuitions at the root of each of these approaches. Our present institutional arrangements can be thought of as an imperfect house to which we are confined. It keeps us warm most of the time and provides much of what we need from a house, but it also has many noticeable flaws that reduce our quality of life—the roof leaks, the heater often doesn't work, there are structural problems, etc. In such a situation, blueprints for an ideal house that solved all of these problems would not be particularly useful—especially because tearing down the house and building a new one is not an option. In the ideal house the roof wouldn't leak, but a proponent of a comparative approach would argue that we don't need to know what the perfect roof is like to know that covering up a hole is a good idea. The fact that an ideal roof shouldn't leak is not a necessary truth in itself, but is instead the expression of a basic value that we can understand without blueprints: we don't want to get wet when it rains. Just as we don't need a blueprint for a perfect house to fix a leaky roof, we don't need a set of ideal institutions to know that, say, slavery should be abolished or free speech preserved.

[43] *Ibid.*

Wiens makes the distinction just mentioned (and discussed above in §2.2), between general ideals and ideal theory. The designs of ideal theory, he argues, are merely developments of basic values given a certain idealized (i.e. not fact constrained) conception of the world. That is, ideal theory takes basic moral inputs and creates ideal institutional outputs according to a set of rules and assumptions. These moral inputs, however, can be applied to normative decision-making without the added step of deriving the institutional outputs of ideal theory. For Wiens, such outputs are entirely superfluous to the task of nonideal theory, which can make use of basic values directly in diagnosing actual injustices "without having to take on board the baggage of ideal theory."[44]

Despite the strength of such comparative accounts, accepting that ideal theory is neither necessary nor sufficient for making the comparative judgments that nonideal theory requires is not in any sense a refutation of its value. When fishing, one really only requires a hook at the end of the line to catch fish. Bait, then, is not necessary and it is certainly not sufficient for the task. It is, however, obviously useful. Similarly, Sen and Wiens succeed in demonstrating that ideal theory is not *needed*, but if it can be shown to assist the progress of justice in the world, why wouldn't it be used and valued? Despite descriptions of ideal theory as "overwrought" or as extra "baggage,"[45] in evaluating and implementing issues of social justice it would be difficult to accuse anyone of considering *too many* perspectives. It

[44] *Ibid.*
[45] *Ibid.*

is the usefulness of ideal theory, rather than any necessity or sufficiency, that I argue for in the chapters that follow.

2.4: Maximization and Calibration

One issue that must be addressed in a discussion of these contrary approaches to the promotion of justice is the question of what exactly justice *is* in each of these views. Justice is a term with many different connotations; one might wonder if they are even talking about the same thing. In this vein, one difference between the comparative and ideal approaches that cannot be overlooked is a divergence in the method of measuring justice that effectually changes the very nature of the thing being measured. These different methods I call *maximization* and *calibration.*

A comparative approach to justice, by rejecting an ideal standard, turns it into an independent and unbounded quantity present in a given arrangement. Drawing on the terms developed in the account of Rawlsian justice from §1, comparative justice is neither *binary* nor is it *systematic*. Regarding the former, when justice becomes an independent quantity the binary distinction between just and unjust becomes as relative as the distinction between long and short. In this understanding of justice it is, like height in Sen's mountain example, something that one can have *more* or *less* of. In the context of such an understanding, the idea that a comparison of two alternatives does not require appeal to a third appears rather intuitive for reasons mentioned in the previous section.

As for the idea of *systematic* justice, a comparative approach like Wiens's in which individual institutional failures can be evaluated in isolation necessarily accepts the theoretical desirability of a piecemeal approach to remedying injustice. Implicit in this is the belief that a positive change in one social institution is a positive change to society as a whole. This divergence from an ideal understanding of the just reveals that what differs between the two approaches is not merely the way justice is measured, but what justice itself *is.* Strictly comparative evaluation, in rejecting an explicit standard, ties justice instead to the idea of preference ranking. The just option and the option preferred by those making the immediate judgment become one.

Such an approach falls under the broad umbrella of what Rawls has in mind in his rejection of what he calls "intuitionism." As he describes it, intuitionism is "the doctrine that there is an irreducible family of first principles [i.e. basic moral inputs] which have to be weighed against one another by asking ourselves which balance, in our considered judgment, is the most just." Importantly, there is "no explicit method…for weighing these principles against one another: we are simply to strike a balance by intuition."[46] An intuitionist, faced with the problem presented earlier of having to choose between privacy and free speech, accepts that no objective or explicit ranking of the two can be made. When the choice must be made there is no priority ranking of rights that can be appealed to, nor is there an institutional ideal that can illuminate the better choice through comparison. The only option is to

[46] *ToJ* p. 34

make a comparative judgment based on one's present intuition as to which alternative is more just.

An ideal approach to justice, on the other hand, does not separate justice and the just society. That is, perfectly just institutional arrangements do not merely embody the greatest amount of some independent quality called 'justice,' they actually *are* justice. They define it. In Rawls's words "the nature and aims of a perfectly just society is the fundamental part of the theory of justice."[47] The idea of being more just than the ideal is meaningless, because justice is not an independent quality. Rather than more or less just, a state of affairs in this approach can be described as *closer* to or *further* from justice (not just in a comparative, but also in a transitional sense). There is, of course, room for intuitive judgment in this ideal understanding of justice.[48] But intuitive judgment about which alternative most closely resembles (or is most likely to allow for the realization of) an institutional ideal is quite different from the more complete intuitionism of relying exclusively on comparative judgments.

The difference between these two conceptions of justice can be understood as analogous to the difference between javelin throwing and archery. In the former, further is always better. Not only that, but the measurement of success (distance) exists independently of the activity of the throw. As a result, there is no absolute "far" or "near"—no consistent standard of success beyond the comparison of the distance of two throws. Archery, on the other hand, measures success by proximity

[47] *Ibid.* p. 9
[48] Recall from §1.2 that Rawls notes that "the measure of departures from the ideal is left importantly to intuition." *ToJ* p. 246

to the center of a target. In this regard there is a limit (a perfect bull's-eye) beyond which the idea of being "more accurate" is meaningless. Accuracy is defined as hitting the center of the target, and this measurement exists only in the context of there being a target. The success of any shot must, as a result, be evaluated with reference to the center.

With this fundamental difference between comparative and ideal understandings of justice in mind, I move now to an examination of a further division of ideal guidance approaches into two types: linear and nonlinear.

2.5: Linear and Nonlinear Transitional Theory

The idea of ideal guidance in nonideal circumstances is often presented unflatteringly as an attempt to prescribe actions that will make a set of institutions more closely resemble an ideal arrangement immediately and as much as possible.[49] Such an approach would necessarily be based on the idea that a single linear ranking of social arrangements is possible with ideal justice at one extreme. Thus, an argument against this approach, the *linear transitional* method mentioned in the discussion of Rawls in §1, is essentially an argument against the *sufficiency* of ideal theory in making comparative judgments. As noted earlier, linear rankings of this type are extremely difficult, if not impossible, to make. The linear transitional approach would also, as Sen rightly points out, create serious problems with

[49] "A transcendental approach cannot, on its own, address questions about advancing justice and compare alternative proposals for having a more just society, short of proposing a radical jump to a perfectly just world." Sen (2006) p. 218

evaluating the desirability of "second best" arrangements when the ideal is not possible. He illustrates the problem by analogy: "a person who prefers red wine to white may prefer either to a mixture of the two, even though the mixture is, in an obvious descriptive sense, closer to the preferred red wine than pure white wine would be."[50] That is, in some (even many) cases an arrangement that more closely resembles an ideal than the current one (assuming similarity could be consistently judged) may actually be less desirable overall. This problem of second-best solutions has major implications for normative theorizing as a whole, across disciplines, and will be the focus of chapter 2.

Although the problems with a linear application of ideal theory are legitimate, there has been relatively little exploration of the idea that the linear transitional model might *not* exhaust the possible methods of applying ideal theory to nonideal circumstances. The alternative to this method is what I call the *nonlinear transitional* application of ideal theory. While both of these transitional models have as their goal the ultimate realization of an ideal institutional arrangement, the nonlinear approach takes a broader and more long-term view of progress toward this goal. While a linear application values actions that make institutional arrangements more *similar* to an ideal, a nonlinear approach instead values actions that make a society *more capable of eventually achieving an ideal*. Such capability is judged neither in terms of similarity to the ideal nor in terms of immediate desirability, and

[50] Amartya Sen, *The Idea of Justice* (Cambridge, Massachusetts: Harvard University Press, 2009) p. 16

may in fact recommend decreases in both in the name of long term realization of the ideal.

To stretch Sen's wine analogy to the limit of its applicability: when someone with a glass of white wine is presented with an array of alternatives (but none of them his ideal of red), the linear method would choose an unpleasant red-white mixture and the comparative method would choose something that tasted better than white. Meanwhile a nonlinear method, recognizing that the choice was not an isolated event but a series of such events, would choose what ever option made it most likely that future choices would include red wine. Such a choice might very well taste worse and be more dissimilar to the ideal than the original white. That is, a nonlinear transitional mode of thinking may actually lead to the selection of an arrangement that appears *worse* in the short term for the sake of an ability to approach the ideal in the long term.

Despite some characterizations of ideal guidance as futilely linear, it is this nonlinear approach that not only *should* be considered the proper method of non-ideal application of ideal theory, but actually has been. A. John Simmons, in his recent insightful exploration of the nature of the ideal-nonideal distinction in Rawls's work, engages the issue of transitional justice and plants the conceptual seed from which much of this essay grows. He argues that Rawls himself did not view the perfectly just institutions of ideal theory as requiring that nonideal options be ranked according to immediate similarity to the ideal. As he puts it, "if it is ne-

cessary to take one step backward in order to take two steps forward, Rawlsian nonideal theory will endorse that step 'away from' resemblance to the ideal."[51]

The transitional nature of this application of ideal theory highlights what I consider a key point of divergence between comparative and ideal approaches to justice. It is very likely that much of the time both approaches will lead to identical policy decisions. In decisions about things like slavery, under most circumstances, both a strictly comparative and an ideal approach will lead to a policy of abolition. The comparative approach makes a judgment on the basis of the inherent moral unacceptability of slavery, the ideal on the basis of an understanding that abolishing slavery will allow society to make long term progress toward fully just institutions—but the output is the same in both cases. However, the willingness of non-linear approaches to take "one step backward" for the sake of future progress can lead to decisions that comparative theory, especially failure analysis, could never recommend. If, as Rawls notes, "there may be transition cases where enslavement is better than current practice,"[52] then a choice would have to be made; a strictly comparative theorist would have to defend, in some cases, a policy that was undeniably more desirable in the short term, but set back progress towards comprehensive

[51] Simmons (2010) p. 23

[52] *ToJ* p. 248: "suppose that city-states that previously have not taken prisoners of war but have always put captives to death agree by treaty to hold prisoners as slaves instead. Although we cannot allow the institution of slavery on the grounds that the greater gains of some outweigh the losses to others, it may be that under these conditions. . .this form of slavery is less unjust than present custom," because "in time it will presumably be abandoned altogether, since the exchange of prisoners of war is a still more desirable agreement." That is, institutionalizing slavery may appear to move society further from the ideal of, presumably, equal liberty for all; but this development is desirable from a transitional perspective as it may allow for the development of arrangements that are more just than the status quo.

justice in the long term.[53] To adopt a comparative approach appealing only to basic values and available choices would necessitate a dismissal of the potentially limiting consequences of immediate decisions on long term paths toward justice. Going against such a myopic view in order to condone or implement an unjust or undesirable policy in the short term for the sake of long-term gains in justice would necessarily imply some conception of an ideal arrangement in the name of which present sacrifices could be made.

A criticism of this nonlinear transitionalism can be made, of course, on epistemological grounds. One might object that we simply don't have the predictive ability to know what the long term consequences of present actions will be, and as a result should limit ourselves to options known to be immediately feasible. Any claims about a long term trajectory toward comprehensive social justice, it could be argued, are guesses at best. This is an undeniably valid point; situations of complete ignorance regarding future consequences of present decisions are more than just possible, they are likely to make up the majority of the choices we face. In such situations, Simmons observes:

> it may seem acceptable to cross our fingers and just accept whatever comparative gains in justice we can get or single-mindedly attack some particular, salient injustice. But it is important to see that committing ourselves to such practices as a general rule would in fact

[53] As Simmons observes: "Where 'comparative gains' or targeted attacks in fact set back the cause of overall social justice, it is hard to see why anyone who is committed to that cause would regard this as nonetheless a positive development." Simmons (2010) p. 24

amount to an abandonment of the goal of any systematic theoretical guidance of political practice.[54]

There are inevitably many situations in which we will lack the predictive power required for nonlinear transitional judgments that condone or implement visibly unjust or unpleasant policies for the sake of long term progress toward justice. But accepting purely comparative judgments as a frequent necessity is very different from accepting such a process as a rule. That is, the epistemological confidence required for nonlinear judgments may rarely be achievable; but in situations where we can say, with some acceptable level of confidence, that a particular option which is the most desirable comparatively will create long term transitional limitations if chosen, a choice will have to be made between the two approaches. To treat purely comparative evaluation as a rule would require advocating a policy that had known long-term transitional limitations. A rejection of the possibility of ideal guidance would, in fact, rule out such transitional considerations completely.

A second important strand of criticism regarding the nature of nonlinear judgments is that accepting them as a valid part of the decision-making process can open the door to very dangerous political abuse. Political action, especially in democratic regimes where voters must at least theoretically approve of policy choices, must normally appeal to the idea of making society comparatively more just than the status quo, or at the very least improving it in some way. The ability to enact

[54] *Ibid.*

patently unjust legislation in the name of future transitional gains would free decision makers from immediate accountability to the well-being of the population. In fact, one need not look far to find examples of precisely this kind of abuse in the world. Political arguments for the shrinking of governmental activity and the dismantling of welfare programs (euphemistically referred to as "austerity") are based precisely on the notion that giving up the tangible benefits of government action and intervention now will stimulate economic growth in the long term to the benefit of everyone.

But putting aside arguments about neoliberalism and questions of whether such transitional austerity policies are proposed in good faith, it must be noted that the uncertainty of popular choice in situations of potential long-term gain is enough to raise concerns over the nature of such transitional judgments in a democracy. It cannot be assumed that informed democratic citizens will be willing to vote against immediate benefits to themselves for the sake of possible long term gain—especially if such social gain is realized not merely *later* in life, but *after* they have died. Thus, nonlinear judgments, made in good faith or otherwise, may well conflict with democratic judgments.

Lastly, in apparent conflict with Rawls's explicitly Kantian roots, nonlinear transitional judgments may be accused of appealing to utilitarian justifications.[55] The realization of fully just social arrangements for people in the future and, one

[55] *ToJ* pp. 179-181: "…the principles of justice manifest in the basic structure of society men's desire to treat one another not as means only but as ends in themselves." On the other hand "utilitarianism does not regard persons as ends in themselves." This idea of treating individuals as "ends in themselves" is probably the most famous aspect of Kant's moral philosophy. See: Immanuel Kant, *Groundwork of the Metaphysics of Morals.*

might presume, the aggregate utility gained through such a realization, appears to be used as a justification for present injustice and even suffering. This leads then to the classic utilitarian problem of justifying the suffering of a few for the greater benefit of many. An industrial worker in the 19th century would find little consolation in the fact that his suffering is contributing to the economic development necessary for progress toward broader economic justice one hundred years in the future.

This apparent tension between nonlinear transitional judgments and a commitment to the inviolability of individuals as ends in themselves must be resolved if such judgments are to be compatible with a liberal theory of justice. This issue, as well as the preceding question of democratic compatibility, will be taken up in chapter 3.

3. Utilizing Ideal Theory

In this introductory chapter I have sketched an outline of the core issues that mark the divide between *ideal* and strictly *comparative* approaches to promoting just outcomes and institutions through political decisions. The key points of divergence are, to recap briefly:

a) A comparative approach evaluates options without appealing to an ideal arrangement. It seeks to rank immediate alternatives relative to one another in order to find the most desirable available option. An ideal guidance approach evaluates options by taking into account, in some way, an ideal arrangement toward which the society should strive.

b) A comparative approach appeals only to basic moral values for guiding action. An ideal approach appeals to specific principles or institutional arrangements.

c) A comparative approach treats gains in social justice as an issue of *maximization*, treating justice as an independent and unbounded quality like height. An ideal approach treats the promotion of justice as an issue of *calibration*, in which justice is defined by and evaluated with an eye to the ideal state, rather than being an independent quality.

With these differences established, recall also that an argument for the possibility of using ideal theory to improve our nonideal decision-making must proceed with certain constraints in mind:

1. Ideal theory is not directly applicable. Its value comes not from dictating individual policy choices without mediation, but from providing valuable insights that contribute to the intelligent application of fact constrained nonideal theorizing.

2. Ideal theory is not to be used for the *linear* evaluation of alternatives. The usefulness of ideal theory lies in its application to long-term, carefully considered, and transitional approaches to justice, rather than the pursuit of immediate similarity to an ideal.

3. Normative ideals can exist independently of ideal theory and, as a result, the value of these ideals does not justify the value of ideal theory. An argument for the importance of ideal guidance must be derived not from these basic normative inputs, but from the practical utility of an institutional ideal.

4. The value of ideal theory is not the same in all cases, but rather depends on nonideal circumstances. While it may provide clear guidance for choice in some situations, in others it may merely provide useful information for choosing amongst a set of options in which none are clearly better than others in the short or long term.

In light of these basic parameters, I proceed in the remainder of this essay to present two important aspects of the way in which idealized conceptions of fully just institutions can provide valuable insights for the development of fact constrained and fully contextual plans of action.

First, in the next chapter, I examine the idea of "second-best solutions" as they apply to theorizing about justice. The problem of second-best solutions is occa-

sionally mentioned in recent literature, but rarely analyzed with the depth that it deserves. I take a closer look at the implications and limitations of the idea that deviation from the complete realization of a set of conditions makes finding a second best alternative much more difficult than one might initially suppose. I argue that the limitations of the "general theory of second best" certainly apply to ideal theorizing, but examine also the equally important and frequently overlooked negative corollary within the theory which creates serious problems for strictly comparative theorizing as well. Given the inescapable limitations on second-best solutions in all areas, I present an argument that these limitations are accounted for more effectively through ideal guidance than through myopic comparative decision making.

Second, in the third chapter, I look at the role of future path dependent outcomes on the process of present institutional design and development. This process, intimately connected with the idea of a nonlinear transitional approach, deals with the possibility of predicting future paths of development toward a fully just basic structure both (a) in predicting the most effective routes toward justice and (b) in identifying paths that, although desirable in the short term, limit or prevent the long term development of fully just institutions. This discussion is intertwined with an examination of the idea of feasibility and the methods through which we might be able to judge what we as a society are capable of doing in the future, beyond immediate constraints and considerations.

The argument developed in the remainder of this essay should by no means be considered an exhaustive account of the ways in which ideal theorizing can improve our judgments about the institutional choices we face. I believe, however, that the ideas presented in the following chapters are sufficient first to defend the possibility of ideal guidance against recent attacks, and second to move considerations of ideal theory toward new ways of thinking about bridging the gap between political theory and political practice.

Chapter II
Second-Best Solutions

> *If it is not necessarily good to increase the size of free trade areas, if markedly deleterious consequences can follow from an increase in the number of industries which act competitively, in sum, if no good may be accomplished by behaving every day in every way better and better, how much more treacherous becomes the task of him that would offer cogent economic advice.*
>
> William J. Baumol[56]

With the necessary conceptual foundations laid in the first portion of the essay, I turn now to the real work of justifying my central thesis: that the realistic utopias of ideal theorizing *can* and *should* be utilized in evaluating immediate policy alternatives. While the first chapter sought primarily to establish what exactly ideal theories of justice *are*, the question now at hand is what we can *do* with them.

This second chapter begins to answer that question through a development of the concept of "second-best" optima. Although the limitations of second-best solutions are often invoked to dismiss the possibility of ideal guidance, robust analysis of the issue as it relates to social justice is almost nonexistent.[57] To remedy this, I

[56] William J. Baumol, "Informed Judgment, Rigorous Theory, and Public Policy," *Southern Economic Journal* 2 (1965): pp. 137-145

[57] Robert Goodin's engagement with the issue is probably the most complete account, but even he spends only a few pages on it. See: "Political Ideals and Political Practice," *British Journal of Political Science* 25 (1995): 37-56

begin in §1 with an examination of the first general formulation of this issue, dating back to the mid-twentieth century, from which most modern analysis is derived. In doing so I develop an account that is more complete and nuanced than the generally brief references in recent literature. The theoretical implications drawn from this account extend much further than is often supposed. §2 addresses the apparent impossibility of non-intuitive evaluations of immediate policy alternatives, but also points toward a way of getting around this impossibility problem. §3 looks at the ways in which strictly comparative decision-making fails to overcome this problem. §4, then, looks at how an ideal guidance approach can avoid, to some extent, the problems of second-best solutions by using a nonlinear transitional method of evaluation. In the final section I take a step back to gain perspective and reflect on the fact that despite the benefits of such an approach, there are still deep uncertainties that we may never escape.

1. The General Theory of Second Best[58]

In 1956 economists R. G. Lipsey and Kelvin Lancaster published an article titled "The General Theory of Second Best" and in doing so captured the essence of

[58] For the sake of clarity, I would like to note here that there seems to be no set convention for hyphenating "second best." I often add a hyphen to avoid ambiguity when the term is used as a descriptor, e.g. "second-best solutions"

an issue that pervades not just economics, but normative social science as a whole; the problem had admittedly been observed in various individual cases for years, but had until then lacked a general expression to bring together such cases under the umbrella of a single phenomenon.[59] The core of the problem they observed was this: for a given optimal situation in which a set of interdependent conditions are all met, if in practice one or more of those conditions are not attainable then the *second* best alternative is *not* necessarily the realization of as many of the original conditions as possible.

As a general example: if your ideal is the satisfaction of five conditions but something limits the satisfaction of one of the five, you might initially assume that the second best option must be to satisfy the remaining four. However, this is not necessarily, and in many cases not even likely, to be the case. While Lipsey and Lancaster observed, and demonstrated mathematically, the economic manifestation of the problems of second-best solutions, the core idea of their general theory is not limited to economics or Pareto optimality. In accepting this broad understanding I share the interpretation of the handful of political theorists who have explored the implications of second-best optima in the political sphere. The general theory is on its surface an economic statement, but this is just one manifestation of a fundamental problem for constrained and interdependent optima of any kind.

[59] R.G Lipsey and Kelvin Lancaster, "The General Theory of Second Best," *The Review of Economic Studies* 24 (1956): 11-33

As a preliminary point of terminological clarification: I use the term *condition* to denote any relevant factor in the evaluation of a state of affairs. Conditions may be satisfied, constrained, or unsatisfied.

1) *Satisfaction* of a condition may take various forms including the presence or absence of something, the maximization or minimization of some quality, or the passing of some necessary threshold. In short, satisfaction of a condition is the ideal state of the condition in which it can be considered to be "met" *given that the other conditions are also met* (more on this in a moment).

2) A *constrained* condition is, for some permanent or eventually remediable reason, unable to be satisfied in accordance with the optimal state.

3) *Unsatisfied* conditions *could* be satisfied in terms of what it would take for them to be met in optimal conditions, but are not. This is due to the lack of desirability of such satisfaction in suboptimal conditions (i.e. with the existence of at least one constraint.)

The foundation of the problem of second-best solutions is the idea of *interdependence*. When, say, the maximization of a set of functions is considered ideal, one might initially make the assumption that movement toward the ideal within each individual function will always, like basic economic commodities, have positive marginal utility. While this can be the case if each function is completely independent of the others, when the desirability of a state *x* of a condition is contingent upon the state of other conditions, a constraint upon one of those other conditions may make achieving *x* undesirable. Progress toward *x*, then, would not

necessarily be a good thing. As a result of this problem, a departure from the ideal in one condition may result in a second best solution in which *all* of the other conditions are also unsatisfied despite the possibility of their satisfaction (that is, what their satisfaction would be in optimal conditions).

Although this generic formulation may at first seem counterintuitive, I present now a few everyday examples which reveal that it is actually quite simple. To borrow Robert Goodin's example,[60] suppose you are buying a car and you would ideally prefer a new silver Rolls Royce. If there are no cars on the lot that meet all three of these conditions (new, silver, and Rolls Royce) your second choice may well be a car that deviates from *all* of the initial conditions rather than a car that satisfies two of them. That is, you might prefer a one-year-old black Mercedes (which satisfies none of the ideal conditions), to a new silver Ford (which satisfies two of them). Simply put, there is no way of knowing *a priori* what the second best option is when the optimal conditions cannot *all* be met.

The nature of interdependence in suboptimal situations can be observed from a different angle if, say, you want cookies and milk but have only cola to drink. When one optimal condition is constrained (no milk), you might find that you would rather have some other food with the cola (intentionally leaving the second optimal condition unsatisfied as well), or that you would rather not have anything to drink with your cookies (moving even *further* away from ideal in the drink condition), or that you would rather just not eat anything at all. When the goal is to satisfy a set of

[60] Goodin (1995) p. 53

related conditions, departure from one may make the rest wholly or partially undesirable.

Although applied initially and most frequently in the realm of economics, and in particular to Pareto optimality over a set of functions, the generality of the theory of second best (which I will refer to also simply as the 'general theory') allows for its extension far beyond economics. Any ideal that takes into account multiple inter-dependent optima will find it difficult to escape its limitations. Thus, the problem of second-best solutions has profound implications for political decision-making. Decisions regarding social justice in an imperfect world must necessarily choose between arrangements in which some, and frequently all, of the conditions of a fully just society remain heavily constrained.

A simple institutional example of this concept that is relevant to the current discussion is the idea of designing rules or institutions under the (untrue) assumption of full individual compliance. Any attempt to coordinate activities based on the "honor system" demonstrates the risks of implementing a set of policies that is ideal when everyone complies, but may be undesirable if the compliance condition is constrained. For example, many retail stores must increase the price of certain goods for everyone in order to offset their losses due to theft. This is, however, a necessary second best alternative. The ideal arrangement in this situation would clearly be that (a) everyone complied with the idea that one should not steal and (b) the store had lower prices for everyone because it did not have to compensate for loss due to theft. If, however, there is a significant lack of compli-

ance with the "no stealing" condition then the second condition, (b), is no longer desirable. Maintaining lower prices that do not account for theft in such a situation might cause the store lose profitability to the point that it goes out of business, depriving the owner and employees of jobs and the greater community of access to certain goods. Thus, (a) and (b) form an interdependent set of optimal conditions, with the desirability of one dependent on the presence of the other.

The basic problem of second best alternatives can, of course, be scaled up to characterize national institutions and raises interesting questions about the way we think of the institutions that govern our lives. It is widely accepted that democratic governance is the ideal system for a society of informed voters. But if this second condition (that voters are informed) is not met, are democratic institutions still the most desirable? In other words, to what degree are democracy and an informed populace interdependent optima? If, rather than voting based on their considered interests, a majority of the population votes for the candidate that has the largest advertising budget or seems most like someone they would "have a beer with," is a democratic system sans informed voters actually the second best option? The process of electing legislative representatives that are presumably more informed already accounts for this to some extent. But how far should the idea of representation go? These questions are, of course, incredibly complex and contentious and I will certainly not try to answer them here. Hopefully, however, they have provided a rough illustration of how difficult questions of second best solutions

under constrained conditions can be, especially as the number of considered conditions grows.

This understanding of the problem of second-best solutions is an important part of many recent critiques of the practicality of ideal guidance.[61] Even if a comprehensive ideal of just institutions is sound, the argument goes, knowing the ideal conditions does not tell us what the next best nonideal choices are. Given the inherent limitations of constrained conditions, a departure from even one of the ideal conditions may well require that we depart from all of the other conditions as well. In such cases, ideal theory would offer no guidance; we could not know that any of its conditions remained desirable under constraints in which they could not all be simultaneously met. This is the problem that underlies Sen's claim that ideal theories of justice cannot "address questions about advancing justice…short of proposing a radical jump to a perfectly just world."[62] This critique has undeniable merit. What holds under ideal conditions of simultaneous satisfaction of all conditions does not necessarily hold given the presence of constraints.

The preceding account of the implications of the general theory is about as far as most recent analysis goes. However, it is not complete. There is, as Lipsey and Lancaster note, an "important negative corollary" in the theory.[63] While it is true on the one hand that deviation from one optimal condition may require that a second best arrangement deviate from all of the others as well (undermining the validity of their guidance), the converse of this problem is equally troubling: given a situation

[61] See: Goodin (1995). Swift (2008). Wiens (2011). Sen (2009)
[62] Sen (2006) p. 218
[63] Lipsey and Lancaster (1956) p. 11

in which many optimal conditions are constrained, the fulfillment of any one condition may not be an overall improvement. In this way, the problem of second-best optima can approached "from two quite different directions," both of which will play important roles in this chapter.[64] (I will refer back to these concepts frequently as (D1) and (D2), so remembering this distinction is of particular importance):

Direction 1 (D1): We may assume the existence of a constraint on one or more of a set of conditions and examine what effect this has on the applicability of the rest of the ideal conditions to a second-best solution. This can be thought of as the broad limitation, dealing with the applicability of all optimal conditions in constrained circumstances.

Direction 2 (D2): We may assume the existence of a large number of constraints and then attempt to identify the effect and desirability of change in any one condition. This can be thought of as the narrow limitation, dealing with the difficulty of evaluating piecemeal changes in any one area in constrained circumstances.

The first approach inquires into the second-best optimum for a *system*, the second into a second-best optimum for a *single condition.* In the realm of social justice the distinction here is subtle but important. While the first problem brings into question the legitimacy of institutional ideals in nonideal circumstances, this

[64] *Ibid.* p. 13

second aspect creates serious problems for the idea that addressing injustice in any one area will lead to improvement overall or that remedying a particular injustice is in itself beneficial. In terms of the above distinction, then, (D1) creates problems for ideal theorizing and (D2) creates problems for nonideal or comparative evaluations. Assuming the goal of a comparative approach is to make society as a whole more just, and not just to move single-mindedly from one instance of institutional failure to the next, it must be taken into account that piecemeal solutions might not be overall improvements.

This broader understanding, then, appears to create a serious problem not just for theories of social justice but also for normative theorizing as a whole. It is not hard to see why one economist described the general theory as "capable of yielding a rich harvest of havoc."[65] And it is, unfortunately, a problem that does not seem to have any real means of resolution. The best that can be hoped for is for an approach that reduces the negative effects of this uncertainty as much as possible. In light of this, I argue in this chapter that while no method of ameliorating injustice is immune to the complications of second best solutions, the proper use of transitional ideal guidance can reduce these complications to a greater degree than otherwise possible with strictly comparative approaches.

[65] Baumol (1965) p. 138

2. Impossibility

Before examining the way in which the limitations of the general theory play out in comparative and ideal approaches it must be noted that, contrary to seeming insurmountability of the problems raised, its authors "did not intend their general theory of second best to be interpreted as an impossibility theorem."[66] While some may take the theory as a demonstration that "one can say nothing in the absence of universal optimization,"[67] the constraints of the real world do not ultimately leave us *completely* in the dark. Faced with constraints on optimal conditions, the general theory does not imply that second-best solution(s)[68] do not exist or that they are unknowable. They must appeal instead, as Morrison astutely notes, to "a different *type* of maximum."[69] That is, Pareto optimality may remain the first-best choice, but the dependence of this optimal state on the simultaneous realization of all of its conditions means that evaluation of constrained alternatives cannot be based on direct similarity to optimal conditions. But the evaluation of second-best solutions is still *possible.* In order get around the impossibility of linear comparison, a useful method of evaluation must appeal to a *fundamentally different measure of what constitutes the "best" amongst nonideal alternatives.*

---

[66] Clarence Morrison, "The Nature of Second Best," *Southern Economic Journal* 32 (1965): 49-52
[67] E.J. Mishan, "Second Thoughts on Second Best," *Oxford Economic Papers* 14 (1962): 205-217
[68] Lipsey and Lancaster (1956): "it is important to note that…there will be a multiplicity of second best optimum positions. This is so because there are many possible combinations of constraints with a second best solution for each combination." p. 13
[69] Morrison (1965) p. 50. Emphasis mine.

The appropriateness of the application of these ideas to social justice may require some clarification. In the Pareto optimal conditions under consideration in the economic study of second best solutions, each individual function is what I call a *conditional good*. That is, the realization of each individual function can be said to be a good thing with certainty only in the context of the realization of all of the other functions. Without such simultaneous realization, the desirability of progress in any one function cannot be known *a priori.*

The nature of the conditional good in this context has clear parallels to the nature of the various conditions of an institutional ideal. Majoritarian democracy can be good *if* there are constitutional protections for minority groups. Some level of income inequality is good *if* competition for higher paying positions is open to all. The conditional nature of many social goods, then, means that when certain conditions are unsatisfied, the evaluation of a second-best arrangement must appeal to a metric other than similarity to the ideal; even if it was possible to make linear comparisons of groups of social institutions, they would be invalid. But as I mentioned above, this does not make meaningful evaluation impossible. The ability to address this need for an alternative metric under nonideal conditions is one of the primary advantages of the nonlinear transitional methodology, which will be explained further in §4 of this chapter.

The inability to evaluate nonideal policy based on an idealized set of conditional goods has also led to an alternate conclusion: that such ideal systems have no role to play in real world policy decisions. In their place a comparative approach

would dismiss entirely the usefulness of ideally optimal, but presently constrained, arrangements in making comparative judgments. With first-best solutions off the table, the modified metric of justice in these comparative methods is the evaluation of directly achievable alternatives on the basis of present moral judgments and social preferences.[70] Such a response implicitly accepts that the general theory is an impossibility theorem for (D1), but ignores the implications of the unavoidable negative corollary (D2).

However, the qualities that invalidate direct comparison to ideal conditions when constraints are introduced *also* make it difficult to evaluate the benefits of progress in any one area in the context of a large number of constraints. (D1) and (D2) are not different problems; they are different aspects of the same problem, and as such must *both* be addressed. A theory that abandons any comprehensive evaluation of ideal conditions on account of the limitations introduced by the general theory must also justify the basis on which individual attempts to remedy injustice are believed to be overall or long-term improvements. That is, it must be shown that the general theory can be an impossibility theorem in (D1) but *not* in (D2).

[70] An analysis of social preferences and the limits of collective decision-making in comparative evaluations such as these can be found in Amartya Sen's insightful work on social choice theory. See: *Rationality and Freedom* (Cambridge, Massachusetts: Harvard University Press, 2004)

3. Comparative Measures of Second Best

Responding directly to Goodin's explication of the limitations inherent in evaluating second-best solutions, Swift captures the essence of the comparative approach in his description of the proper foundations of policy in nonideal circumstances as "a sensible refusal to fetishize 'ideals' or 'principles,' a thoughtful evaluation and weighing of the different fundamental values at stake, and a social scientifically informed all-things-considered judgment about which options within the feasible set are preferable to others."[71] As I've hopefully made clear by this point, the defining feature of the various comparative approaches to justice is the rejection of ideal conceptions of fully just institutions—the refusal to 'fetishize' the abstract just society.

But the vacuum left behind by a refusal to rank present policy alternatives based on an ideal institutional standard or goal must be filled by something. The alternative method of evaluation presented is generally an appeal to loosely defined moral standards (e.g. "fundamental values,"[72] "ordinary moral reasoning"[73]). How we are to decide what the appropriate moral standards are remains for the most part unaddressed. I do not, however, bring up these points as criticisms. The methods for deriving appropriate moral standards in a given comparative evaluation are distinct from the methods of *using* those moral standards. Throughout this essay I myself am arguing for the *usefulness* of an ideal theory of justice, rather than

[71] Swift (2008) p. 377
[72] Goodin (1995) p. 56
[73] Wiens (2011) p. 23

for any specific method of deriving its *content*. However, even if the question of how moral standards are decided upon is left unanswered, the way in which these moral standards might be used can be closely examined.

There are, it appears to me, two main paths that a comparative approach can follow in utilizing moral (rather than institutional) standards in evaluating feasible alternatives. Both fall broadly under the umbrella of *intuitionism* discussed in the first chapter, but can be distinguished further. The first comparative method I call *comprehensive* moral guidance. This position seeks to identify the essential moral considerations that justice requires without attempting to describe the institutions that would embody those ideals. For example: liberty of speech, motion, and religious belief; equality of basic rights; access to basic sustenance or healthcare; freedom to play and create and flourish as a human being. These moral values are basic in that they describe goals without institutions. In this sense they are more flexible than ideal institutional prescriptions; they are adaptable to specific circumstances in which institutional reform might not be possible. As Goodin puts it, after rigid ideal institutions are put aside, "timeless truths, ideally ideal ideals, remain. All that has to go are context-free political prescriptions for realizing them."[74]

Replacing ideal institutions with broader conceptions of moral ideals would certainly appear to make nonideal decision-making more flexible, but does it manage to avoid the limitations of constrained second-best theorizing? I argue that it does not. Despite claims that the problem of the inapplicability of ideals "affects

[74] Goodin (1995) p. 56

only particular principles *qua* expressions of values, not the abstract values themselves,"[75] a comprehensive set of moral principles does not escape the problem of interconnectivity that makes them conditional goods. That is, moral principles that are ideal when realized simultaneously are not necessarily beneficial when pursued under constrained conditions. I do not think it necessary to go into too much detail on this point, as the problems of measuring progress relative to a set of ideal conditions that I previously explained apply here in precisely the same way.

I noted above in §2 that (D1) (the broad limitation on second-best solutions) necessitated a different *type* of metric if constrained evaluations were to be made based on ideal conditions. When Pareto optimality is the ideal, second-best alternatives cannot be measured linearly based on similarity to Pareto optimal conditions. In this same manner, nonideal institutional alternatives cannot be compared based on similarity to institutional ideals. While a comprehensive moral framework *does* manage to use a metric other than institutional ideals in the evaluation of institutions, it also changes the nature of the ideal conditions from a set of institutions to a set of moral principles. Thus, a set of moral ideals is just as limited by (D1) as a set of institutional ideals.

The second comparative method can be described as *decisionistic* moral guidance. I use this term with reference to the idea of decisionism. This is, briefly, the idea that a particular moral or legal decision has value not because of any particular content, but because of the legitimacy of the process through which it was

[75] Wiens (2011) p. 12

chosen or of the people doing the choosing.[76] Accepting (D1) as an impossibility theorem for any comprehensive set of optimal conditions, a decisionistic application of moral limitations appeals only to the immediate moral considerations of the parties involved in the decision process. Justice, in this view, is not a specific ideal arrangement; it is whatever those making comparative decisions decide it is. Which of two states is more desirable, or just, is dependent not on any external metric but on the immediate judgments of the individuals involved. It explicitly avoids long term interrelated ideals in favor of a focused evaluation of specific feasible alternatives based on immediate moral judgment. Wiens presents what is probably the clearest formulation of this way of thinking, addressing head-on the frequently ambiguous role of moral judgment in policy decisions. A brief examination of his approach will, I think, allow for a clearer demonstration of the limitations of decisionistic moral evaluation.

In his "institutional failure analysis" approach, Wiens argues for a focus on averting institutional failures rather than looking toward a set of ideals. In this way he is explicitly aligned with Sen's emphasis on remedying *in*justice rather than seeking ideal justice.[77] Wiens, however, actually attempts to provide a specific framework for how such a method might work. In the first step of this process—*identifying* institutional failure—he proposes a comparison of present conditions with feasible alternatives and a ranking of these alternatives based on discussion of

[76] Specifically, I have in mind the works of German political philosopher Carl Schmitt. See: *Political Theology* (1922).
[77] Wiens (2011): "In Sen's words, these judgments identify 'remediable injustices.' On the failure analysis approach, a failure just is a remediable injustice." p.13-14

relevant moral considerations amongst the affected parties. In this stage he acknowledges the problems of individual partiality that ideal arrangements attempt to avoid, but appeals instead to basic universal, or at least widely accepted, moral values upon which comparisons of the desirability of feasible alternatives can be based. In this stage he comes very close to the comprehensive method of moral guidance just discussed, allowing for "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."[78]

In this discussion of the identification of institutional failures he also references (in the typical cursory fashion) the general theory of second best, specifically invoking (D1) in his observation that in constrained circumstances, an ideal principle's "service as an expression of an important value…will be in question."[79] As I explained above, however, when it comes to the moral values that underlie institutional ideals, what's sauce for the goose is sauce for the gander. The nature of interrelated ideals places the same limits on plural moral ideals as it does on institutional ideals.

The more decisionistic aspects of Wiens's thought emerge in the second stage of failure analysis: the diagnosis of failure. The application of moral decisionism is most clearly observed in his claim that "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 cham-

[78] *Ibid.* p. 11
[79] *Ibid.* p. 11-12

pioned."[80] In this position we find an acceptance of (D1) as an impossibility theorem, but no mention of (D2). Each proposal, in this stage of moral diagnosis, "is tentative and experimental, aiming at piecemeal, incremental progress."[81] That is, it rests on the idea that any piecemeal change can be known to be progress which, as has I hope been adequately demonstrated, is not the case in nonideal conditions.

There will most certainly be many times when there simply isn't enough information or time to do anything but make incremental attempts at progress. But rather than accepting the apparent impossibility created by (D1) in all circumstances, I argue next that there are ways around the theoretical paralysis that would seem to result from the full implications of the general theory.

4. Ideal Theoretical Measures of Second Best

Although the alleged inapplicability of institutional ideals in the face of (D1) has been frequently observed, attempts to defend the ideal guidance approach against such claims are almost nonexistent. Simmons specifically mentions the general theory briefly in one paragraph, but despite this brevity he manages to touch on the key to a defense of ideal guidance with his observation that "the idea of 'the second best' is not normally understood to include…the *transitional* aspects of

[80] *Ibid.* p. 18
[81] *Ibid.* p. 22

Rawlsian nonideal theory."[82] It is on this point that the previously mentioned necessity of a different *type* of measurement comes into play. The limitation of (D1) is that given one or more constraints, direct comparisons of similarity to inter-dependent optimal conditions cannot be considered a valid method of evaluating second-best alternatives. If direct comparison to a comprehensive ideal were the only way that such an ideal could be used in evaluating second-best arrangements, then (D1) would indeed be an impossibility theorem. But it is on this point that the crucial distinction between linear and nonlinear transitional theory comes into play.

The linear transitional application of ideal theory, you'll recall, is the principle of evaluating feasible nonideal alternatives based on *immediate similarity to an ideal arrangement*. Clearly, this is precisely the type of approach that is completely invalidated by the implications of the general theory. When all of the ideal conditions are not met, similarity to the ideal state in any or all conditions is not necessarily desirable. A linear approach, then, attempts to evaluate options using the conditionally dependent ideal states as a metric without the conditions that justify those states in the first place (i.e. simultaneous satisfaction). This linear caricature of transitional ideal theory is often held up as an example of its impracticality. There is, however, an alternative.

In the first chapter, nonlinear transitional theory was defined as an approach that valued immediate policy alternatives not based on their similarity to a complete ideal, but on their ability to make a society *more capable of realizing ideally just*

[82] Simmons (2010) p. 25

institutional arrangements. In the present context the significance of this difference is immense. While (D1) may invalidate comparisons to an ideal set of conditional goods as a valid method of ranking immediate choices, the nonlinear approach overcomes this issue by essentially ignoring the linear similarity or dissimilarity of a set of constrained conditions to ideal ones, focusing instead on the likelihood that a policy or arrangement will eventually allow for the simultaneous realization of *all* of the ideal conditions. In other words, progress is not making second-best alternatives more similar to the ideal; progress is the development of the ability to remove the constraints that make second-best rankings necessary in the first place.

While it is true that real world constraints usually create major departures from many ideal conditions, this is in no way undermines that fact that such conditions still represent the first-best situation. The value of similarity to the ideal may be questionable, but the value of actually realizing the ideal is never in doubt. It is only the road there that is thrown into darkness. Keeping this realization as a goal, nonlinear transitional theory puts aside the metric of similarity and looks only at the path to the achievement of first-best conditions. It attacks, rather than immediate injustice, constraints on the possibility of complete justice.

This alternative method of evaluation forms the basis of the Rawlsian idea, mentioned in the first chapter, of being willing to take one step back for the sake of later taking two steps forward in the pursuit of justice. On a scale of direct similarity to the ideal, then, transitional applications of ideal theory may, as Simmons puts it, "dictate the pursuit of 'third-' or 'fourth-best' options instead…if this is what is

necessary if we are ever in the future to actually reach the ideal."[83] Thus the two limitations of linear transitional theory are both overcome by this change in the method of evaluation. First, the impossibility of making objective linear comparisons to an ideal with competing conditions such as basic rights is avoided entirely. A choice between free speech and privacy is no longer a part of the decision process. One of them might conceivably be chosen over the other at some point of course, but such a choice would not be based on the futile task of trying to measure them against each other to see which one brings us closer to ideal conditions. Second, the invalidity of comparisons based on similarity to ideal conditions in constrained circumstances is avoided by changing the metric of progress from *similarity* to *potential*. That is, the potential to eventually and fully realize the ideal, rather than merely resembling it in the short term.

One might at this point notice that such a shift appears to be equally applicable to the comprehensive moral evaluation discussed in §3. If both institutional and moral comprehensive ideals face the same limitations in (D1), shouldn't they *both* be able to retain some of the benefits that ideal guidance provides by shifting metrics from similarity to potential? If basic moral values could respond to (D1) while retaining their greater flexibility, it seems the additional step of specifying principles and institutions in ideal theory would be superfluous. Nonideal policy could be guided by a set of agreed upon moral values just as well, or even better, than it could by an institutional ideal derived from those values. The resolution of

[83] *Ibid.*

this final point is in a sense the climax toward which the preceding discussions have built.

There is one important feature of an institutional ideal that enables it to serve as an evaluative tool without appealing to linear similarity: it has a clearly defined state in which it is "realized." By turning basic values into institutional schemes that embody them, ideal theory creates a goal that is capable of being evaluated as "achieved" or "not yet achieved." That is, it provides a goal based on calibration rather than maximization. A comprehensive set of moral values, on the other hand, does not provide any clear idea of what it means to "achieve" the proper level and balance of each condition. Valuing liberty, equality, human rights, quality of life, etc. is important, to be sure, and necessarily underlies the development of ideal institutions. But in this basic form, moral values do not provide any way out of the epistemological limitations of the general theory. The ability to shift evaluation to capability rather than similarity relies on there being of a state of realization, but what would it mean to describe an arrangement in which liberty and equality are "realized"? Specifying these principles in forms like "freedom of speech" or "liberty of movement," appear to allow for such a possibility, but such formulations are not really basic moral values anymore. Refining and clarifying them in this way turns them into specific principles or rules, i.e. the building blocks of institutions. It is only when such precisely defined principles or institutions are derived from basic values that the limitations of (D1) can be, to some degree, overcome.

5. Conclusion

The difficulty of evaluating second-best solutions cannot be ignored in any attempt to make decisions about socio-political arrangements. Such difficulties are a built-in feature of our ability to envision comprehensive ideals from a constrained standpoint—to look toward what may lie on the opposite edge of the chasm of future uncertainty that stretches always before us. However, as I have argued in this chapter, while a direct comparison of our present condition to an ideal condition cannot provide us with meaningful guidance, an alternative method of comparison may be able to. Rather than utilizing clear or immediate methods of ranking immediate alternatives, such an approach must appeal to admittedly foggy predictions about potential future paths toward justice. However, what I hope to have demonstrated in this chapter is that such clear and immediate methods of ranking do not exist. Confidence in piecemeal comparative progress, as well as *linear* transitional attempts to evaluate feasible options, are both thrown into the darkness of second-best optimization.

A nonlinear attempt to focus on eventually overcoming constraints, rather than settling for blind attempts to optimize second-best alternatives, may offer only a dim light—but it is the best that can be hoped for. Such predictive judgments will often be impossible due to the limitations of our knowledge, and in such cases we may well be forced to "muddle through the best we can…to cross our fingers and

just accept whatever comparative gains in justice we can get,"[84] but there will be times when the choice of whether or not to pursue an immediate comparative gain at the cost of long-term progress toward the realization of fully just institutions will have to be made. It is in these situations that transitional and comparative decision-making diverge, with the former emerging as the more complete method of evaluation.

It is important, however, to note that this transitional approach does not offer any sort of refutation to the claim that ideal guidance is neither necessary nor sufficient for the evaluation of nonideal alternatives. The nature of nonlinear transitional theorizing, and its dependence on prediction, prevents it from providing a precise and unambiguous evaluation of present choices. It can, at best, provide a judgment of some set of alternatives as more likely to allow for the future realization of unconstrained justice. The task of selecting particular political courses of action will, as always, ultimately depend on fully contextual judgments and moral considerations.

With all this in mind, the question of what exactly nonlinear transitional theory *is* in terms of the ideal-nonideal distinction may remain. Nonlinear transitional theory is, in my account, a *partial nonideal theory*. It is a method for deriving guidance from institutional ideals in nonideal decision-making processes, but it is far from sufficient for such considerations. It is merely one factor, albeit a

[84] Simmons (2010) p. 24

very important one, to be taken into account when making a fully informed policy decision.

Questions naturally remain, after an account such as this, about the feasibility of actually removing constraints on first-best alternatives, as well as about the nature of the potentially unjust choices possible through nonlinear transitionalism. I will proceed now to address these issues, among others, in chapter 3.

Chapter III
Path Dependence and Feasibility

> *…there is a question about how the limits of the practicable are discerned and what the conditions of our social world in fact are; the problem here is that the limits of the possible are not given by the actual, for we to a greater or lesser extent change political and social institutions, and much else.*
>
> John Rawls[85]

One natural response to an argument that policy choices should be evaluated in light of their ability to facilitate the eventual realization of fully just arrangements is the thought that, as the old saying goes, "a bird in the hand is worth two in the bush." While the potential rewards may be great, transitional gains in justice are both provisional and uncertain, while comparative gains have, at the very least, results that can be evaluated based on immediate conditions. And while strictly comparative decisions may have unintended consequences in the future, we can at least be somewhat sure that each decision made is beneficial to those immediately concerned. A continuous series of such gains may never result in a perfectly just society, but at least each decision will be comparatively beneficial. Transitional

[85] *LoP* p. 12

predictions, on the other hand, may favor sacrifices now for the sake of potential causal paths that are far from certain.

These concerns are very important and must be addressed if nonlinear transitional theory (hereafter *NT theory*) is to have a firm foundation. It must be shown that the risks associated with attempts to predict potential developmental paths toward justice do *not* outweigh the risks associated with a focus on immediate evaluation. To do this, an examination of the idea of how present choices affect the range of feasible options in the future is essential. Such an examination is the focus of this chapter, which can be divided into two intimately connected strands of thought: *path dependence* and *feasibility.*

Beginning with the former, §1 provides an introduction to the rather intuitive but immensely important idea of path dependent outcomes in social and institutional development. An understanding of the effect of present decisions on future possibilities is absolutely essential to any transitional application of ideal theory. §2 illustrates this importance through an analysis of the concept of *dead ends*—decisions that limit future possibilities in such a way that the realization of a given institutional ideal becomes effectively impossible (or at least very unlikely). §3 then brings together the ideas of path dependence and second-best limitations under the umbrella of NT theory.

The discussion of feasibility begins in §4 with a reexamination of the nature of a *realistic utopia*. §4 and 5 revolve around the central question: what does it mean to say that an ideal is possible or feasible? Or, put differently: what does it mean to

say that an individual or a society has the *ability* to do something? In moving beyond the overly limiting idea that only what is immediately possible can be considered feasible, the idea of feasibility is understood as a process of constantly shifting possibilities over time. The concept takes on a branching structure in which the feasibility of any future state is dependent upon the structure of the connections between possible intermediate states. In §5.3 I take a closer look at the parallel structures of path dependence and feasibility, which mirror each other with a rather aesthetic symmetry. I conclude the section by bringing together the various threads of thought presented in chapters 2 and 3 and weaving them into a unified account of the nature of a nonlinear transitional approach to ideal guidance. This brings my main line of argument to its conclusion. §6 then proceeds to defend NT theory against two important criticisms that might be made against it. Namely, that it is inescapably utilitarian as well as potentially authoritarian. §7 concludes the chapter with an analysis of four necessary limitations on ideal guidance.

1. Path dependence

Path dependence is, at its core, the idea that certain events or decisions affect the range of possible directions in which future development in can proceed. That is, the set of possible options at any given time is shaped and constrained by historical

decisions and events. The decision to take on a large debt, for example, will place serious constraints on fiscal alternatives for years or decades to come. Although often boiled down to the simple idea that 'history matters,' formal study of path dependence is much more than that. That historical events limit the possible trajectories of future development is a claim that hardly requires any justification. However, the recent movement toward a formalized study of path dependence seeks to uncover not merely *when,* but *how* such path dependence occurs. It asks what exactly the mechanisms are that lock in certain decisions and limit future alternatives.

That the choices made in the past affect our options today is obvious, but just how such limitations occur is more obscure. It is in this attempt to formalize path dependence that there is hope for the possibility of using such knowledge to understand not just how yesterday's choices affect today's, but how today's might affect tomorrow's. It should be noted, however, that unlike predictions in the natural sciences, path dependence deals with the immensely complex and unpredictable subject of social and political development. In such circumstances the best that can be hoped for is not a set of claims that "if this happens, this will certainly follow," but instead the somewhat hazier claim that "if this happens, this will (probably) not be able to happen." It is, in this sense, often a negative predictor of what *won't*, rather than what *will* happen.

Scott Page, in a recent essay on the subject, provides a useful survey of some of the mechanisms through which path dependence occurs, which I paraphrase here and illustrate with commonly observable examples[86]:

1) *Increasing returns* is the idea that a certain choice has greater benefits the more frequently it is chosen. Thus, once it is chosen once or a few times, there is an incentive to continue to choose it due to these increasing benefits. Membership clubs at shops or supermarkets often attempt to exploit precisely this mechanism by providing greater savings the more someone shops at that particular store.

2) *Self-reinforcement* is the idea that certain choices, once made, put in place other forces or institutions that then encourage that choice to be made continuously. The decision to lie about something, for example, also creates certain expectations from the people lied to, which in turn encourage the liar to maintain the lie rather than face the social consequences of exposure.

3) *Positive feedbacks* are the increasing benefits of making a certain choice the more other people make the same choice. Although similar to increasing returns, positive feedback differs in that it not only makes further selection of a choice more desirable on its own, it makes it desirable for everyone else that has already chosen. Social networking websites are an example of positive feedback. As more people

[86] Scott E. Page, "Path Dependence," *Quarterly Journal of Political Science* 1 (2006): 87-115

join the same website, its value increases to every individual user already on the website.

4) *Lock-in* is the idea that a certain choice becomes the most desirable choice indefinitely after it has been chosen enough times in the past to go beyond a certain threshold. The QWERTY keyboard is a classic example of this last mechanism. It is certainly not the most efficient layout possible, and in fact slows down typing compared to other layouts. But after it had reached a certain threshold of widespread use, the difficulty of attempting to switch to a different layout became more trouble than it was worth and so QWERTY was locked in as the standard.

Clearly, a formalized path dependence is much more than the simple claim that the past affects the future. It is a very real set of interactions that occur across the social spectrum and that greatly affect the possible range of future choices, giving greater weight to certain choices over others. Although I have presented here a few specific examples of the path dependent mechanisms that constitute a formalized study of the subject, this is primarily to provide some context to keep in the back of one's mind. In the pages to follow I deal with the effects of the existence of path dependencies in general on the transitional development of just institutions, rather than on the specific interactions and methods of attempting to predict path dependencies. An analysis of specific mechanisms would certainly be a necessary

part of any attempts to use NT theory for specific policy evaluations, but it is not necessary for a theoretical examination of the foundations of such an approach.

Progress toward just social arrangements does not occur as a single grand decision—some collective agreement on a new social contract. It happens through innumerable branches of small decisions on the individual and institutional level over the course of decades and centuries. Thus, path dependent processes have significant implications for the pursuit of social justice. If there is any hope of utilizing an institutional ideal to provide some small measure of foresight as we move forward into an uncertain future, it will have to be through an attempt to understand the limitations our decisions today place on our choices tomorrow. An understanding of path dependence is a necessary foundation of any sound transitional application of a theory of justice.

2. Dead Ends

It follows from the basic notion of path dependence that any decision made in the present should, at the very least, consider the potential path dependencies that will result. One particular aspect of these potentialities that I examine now in relation to policy decisions regarding social justice is the potential for *dead ends*. This risk can be explained most clearly with a visual aid:

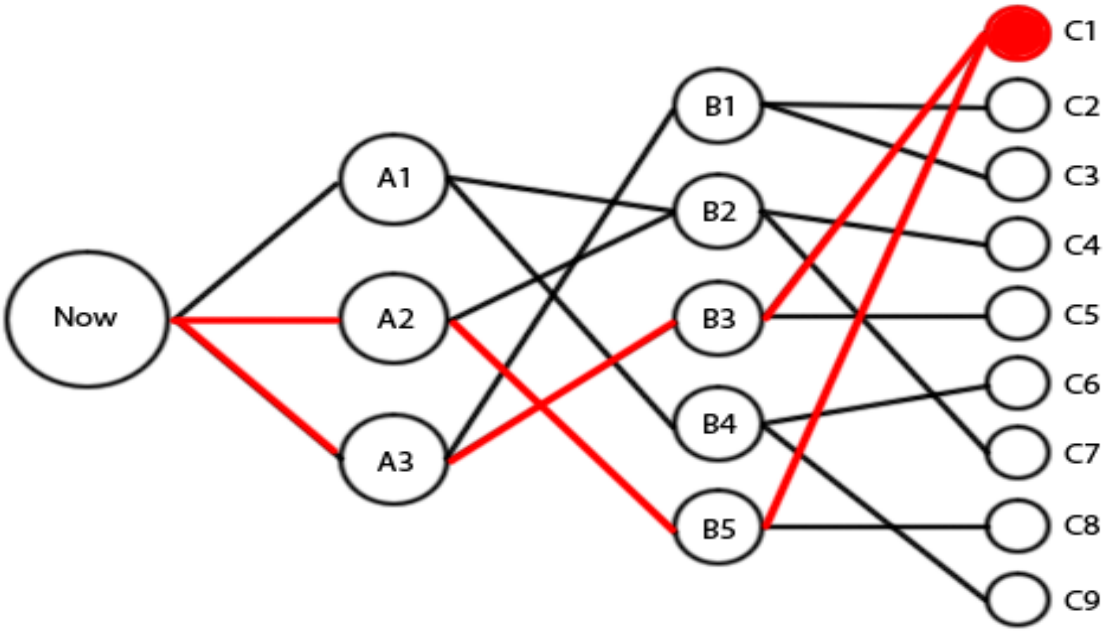

Fig. 2

This path dependence tree (hereafter *PD tree)* represents a series of choices (branches) and the potential outcomes (nodes) that result from each. The leftmost node represents present conditions, and each step rightward is a chronological step forward. The vertical (and thus, numerical) position of each node represents its immediate desirability (that is, without taking into account future paths) relative to all other possibilities at that point in time. For the sake of simplicity, the vertical position of each node represents only ordinal, and not cardinal, desirability.[87] (C1),

[87] That is, it represents only relative rank, rather than an absolute degree of desirability. Although they are spaced equally, the drop in degree of immediate desirability from A1 to A2 is not necessarily the same as the drop from A2 to A3. However, as I will later show, the full potential of PD trees as a method of visualizing path dependent outcomes emerges only when the y-axis is used to measure cardinal desirability.

in red, represents an ideally just society and the red lines denote the two potential paths from the present to this goal. This tree, and PD trees in general, are of course much more precise than real world analysis could ever be—but this example serves well as an illustration of the principles underlying analysis of path dependence.

When presented with such a visualization, one thing that immediately stands out is that *the desirability of a choice in one selection stage is not necessarily correlated with the overall desirability of the alternatives that follow from that choice.* That is, picking the most desirable option in any given round does not mean that the next round of choices will contain overall desirable choices. This can be seen clearly in the possible trajectories from (A1), the most desirable choice in the present. Once chosen, however, (A1) leads not to the most desirable choice in the second round, (B1), but to the second and fourth most desirable choices (B2) and (B4). From these positions the relative quality of available choices drops even further, leading in the next round to, at best, (C4). In this way, the selection of the most comparatively appealing option within any set of choices can lead to comparatively undesirable results over time. Such situations are what I call *dead ends*: trajectories that lead down irreversible paths in which the progression of possible choices either does not improve or gets worse over time relative to the complete set of potential states (including other foregone paths).

The alternative to a method of choosing the most immediately desirable alternative is a transitional selection process that evaluates present options based on possible future paths. With such an approach, (A1) would be off the table due to

the limited desirability of its potential paths and the fact that there are *no* paths from (A1) to (C1), the ideal. Instead, the comparatively less desirable (A2) and (A3) would constitute acceptable options in a transitional sense due to the possibility of continuing on to (C1). Which of these two choices is better is not clear from a purely transitional standpoint as they both lead eventually to the ideal. Some evaluation of the route each takes would have to take into account not just their potential for eventually reaching (C1), but the moral acceptability of each intermediate stage. For example, (A2) might be more immediately desirable than (A3), but if (B5) is a morally unacceptable alternative then the (A3) to (B3) path may be more desirable overall.

Ambiguity of this kind is an unavoidable part of NT (nonlinear transitional) evaluation, and is one of the reasons why I describe it as a primarily negative part of the evaluation of immediate alternatives. In analyzing potential transitional paths to an ideal, NT theory is much more capable of arguing that certain choices should *not* be made due to the potential for dead ends than that certain choices within the transitionally acceptable set should be made over others. This second stage of decision making is, and should be, based on immediate circumstances and on what the relevant parties are willing to bear.

3. Transitionalism and Path Dependence

With the nature of dead ends established, it is necessary now to bring together sections 1 and 2 to examine *how* a conception of an ideal, along with knowledge of the mechanisms of path dependence, can be used to attempt to predict and avoid dead end trajectories. It is important to stress, however, that a developed understanding of the effects of path dependence cannot reasonably be expected to produce any kind of *certainty* about future developments; to do so would be to turn political development into a natural science. The usefulness of ideal guidance, however, does not depend on such unattainable certainty of outcomes. It is instead based on the much more modest claim, established in the second chapter, that even the fuzzy and uncertain predictive evaluations of NT theory are beneficial compared to the apparent darkness through which attempts at second-best evaluations must proceed.

To illustrate the dangers of dead end paths, let us examine now a hypothetical situation in which (1) a clear path dependent mechanism is at work, (2) that mechanism may prevent the long term realization of a complete set of ideals, and (3) avoiding the dead end requires the selection of an alternative that is both immediately less desirable and apparently more distant (in terms of similarity) from ideal arrangements.

Suppose there is a small island nation that is ruled by a violently oppressive and tyrannical military junta. While one very basic institutional ideal for such a

country might be the protection of basic human rights, present conditions are quite distant from the realization of such rights. Even worse, the rapid decay of basic infrastructure due to mismanagement and corruption means that hundreds of thousands are facing starvation. Given the possibility of human suffering on such a massive scale, wealthier nations may well choose to provide monetary or agricultural aid to the blatantly oppressive regime with the hope that the mass starvation of the population might be prevented, at least for a while. As time goes on the regime fails to improve, but aid continues to be delivered for the sake of preserving human life.

This example will, of course, be simplified to an extreme degree compared to real world problems of a similar nature. But nonetheless, this hypothetical scenario can be understood clearly in terms of the three points mentioned on the previous page. First, the continued aid reinforces tyrannical behavior by preventing what would normally be the dire consequences of severe mismanagement, corruption, and violence. Importantly, this effect may get worse with time as the regime becomes more entrenched and less capable of independently supporting its population with every passing year. With each repetition of the decision to supply aid to the oppressive state, then, the consequences of *not* supplying aid next time become more severe. This is a classic example of path dependent behavior, in which each repetition of a decision makes that decision more likely to be chosen in the future.

Second, continued support creates a dead end by reinforcing the continuous selection, via aid to the junta, of the preservation of minimal human rights and well-being over the alternative of allowing major social collapse. Such a choice will be more immediately desirable than the alternative every time, but by reinforcing the conditions that make the choice necessary in the first place it also creates a dead end in which the actual realization of basic human rights becomes less and less attainable as time goes on. That is, the longer aid is given the less likely it is that serious progress will be made toward respecting human rights within the state.

Third, the avoidance of this dead end would require the selection of an alternative arrangement that was both immediately undesirable and more dissimilar to the ideal. If a more complete realization of human rights first required the collapse of the tyrannical regime, then cutting off humanitarian aid might be the only option that made such a realization possible. Such a choice would cause the state to fall into chaos and violence for a while. This outcome is both less immediately desirable (for obvious reasons) and further from the ideal (of the universal preservation of human rights) than the outcome of continuing to supply aid. But unlike the choice to supply aid, it leaves open the possibility of a more just basic structure developing out of the ashes of the old regime, rather than preserving minimally just arrangements indefinitely. And, importantly, the evaluation and selection of such an alternative depends entirely on an analysis of possible future outcomes with respect to the realization of a specific ideal of basic human rights. A PD tree visualization similar to the one in §2 (but slightly more advanced) illustrates

the nature of the branching alternatives and the mechanisms of path dependence in this example:

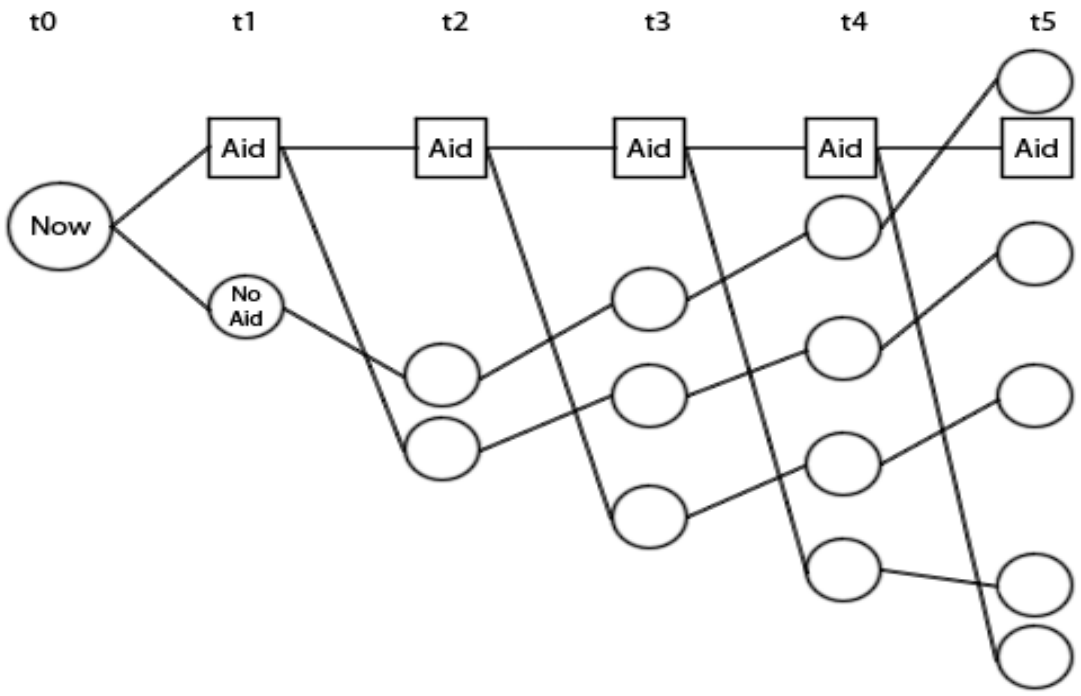

Fig. 3

Unlike the PD tree presented earlier (fig. 2 p. 84), the vertical position of each node here represents not just ordinal rank, but also absolute (that is, cardinal) desirability. Each square represents the choice to supply aid to the oppressed population via the tyrannical government, and thus to preserve the status quo from (t1) onward (hence immediate desirability remains constant). The circles represent cutting off aid and allowing the junta to collapse, with the hope that such a catastrophic failure will allow for the eventual establishment of a more just and

cooperative regime. It is assumed that once the decision to cut off aid is made, the donor state will not change its mind (or, perhaps, governmental collapse will make supplying aid to the suffering population infeasible without great cost) and as a result each node at which this decision could be made results in a single developmental trajectory.

The self-reinforcing nature of the decision to supply aid is visualized here as the fact that with each step forward in time, the decision to cut off aid has more severe consequences. This is presumably because as time goes on the junta becomes more deeply entrenched and, due to its dependence on foreign aid, less capable of organizing the production and distribution of essential resources. Thus every time aid is supplied, it becomes more likely that the same decision will be made next time. The path dependence mechanism at work here is essentially the inverse of increasing returns.[88] Rather than a particular choice becoming more desirable in the future every time it is selected, *not* making that choice becomes *less* desirable with each selection.

At first the undesirable consequences of not supplying aid are less severe and are predicted to lead gradually toward the establishment of an arrangement more desirable than perpetual aid by (t5) at the earliest. As the delay before the cessation of aid grows longer, however, the trajectory of recovery becomes more gradual. Finally, at (t3), the path of continued aid becomes a *dead end*; opting to cut it off at (t3) or later leads to a state of collapse at (t4) from which recovery is no longer

[88] An explanation of increasing returns can be found on p. 81.

possible. While the earlier and less severe collapses led to upward developmental trajectories, the one at (t4) only leads to an even worse state at (t5). The realization of even the modest ideal of basic human rights becomes impossible for the foreseeable future. The choices that remain are (1) perpetual basic subsistence on foreign aid mediated by an oppressive junta, or (2) the anarchy and long-term uncertainty of a failed state with poor recovery prospects.

A method of decision-making that always chose either the most immediately desirable outcome or the outcome that most resembled the ideal arrangement would inevitably follow the dead end path of perpetual aid. An NT approach, however, would advocate a path of action that might be much harsher in the short-term, but that allowed for the future realization of the even better outcomes precluded by the dead end trajectory.

In reality there would, of course, be many more policy options than the two presented here. This entire example is far too simplified to say anything about what should actually be done in a similar scenario in the real world with all its complexity and ambiguity. But hopefully it serves its purpose of illustrating some important aspects of NT theory, dead ends, and path dependent outcomes.

In light of all that has been said so far, the purpose of NT evaluation can be divided into two parts: 1) to draw guidance from an ideal goal despite the inherent limitations of constrained second-best arrangements and 2) to predict transitional dead ends and guide comparative decision-making away from them. Both of these tasks, it should be noted, aim to *enhance* comparative evaluation—not replace it.

That is, NT theory is not a directly applicable output which can rank various courses of action independent of any other factors. Although it is a method of applying ideal theory to nonideal decision-making in a practical and beneficial way, it serves only as a *part* of the evaluation process for a nonideal, fully fact-constrained set of alternatives.

With a more developed notion of path dependence, it becomes clear also that the two tasks mentioned above are inseparably intertwined. While chapter 2 argued for the possibility and value of an alternative metric of evaluation that was able to make use a set of ideal conditions in nonideal circumstances, it is through the idea of path dependent outcomes that this alternative metric finds its expression. An evaluation of path dependent outcomes in pursuit of an ideal state, on the other hand, requires that the possibility of using an ideal as an evaluative tool first be justified. Thus, path dependence and second-best limitations are the two legs upon which NT theorizing stands.

4. The Nature of the Ideal

Throughout this essay I have left the content of ideal theory somewhat vague, attempting instead to explore the usefulness of ideal guidance in general rather than of any specific conception of justice. But in light of the arguments presented so far in

this chapter and in the previous one, a reexamination of the nature of the ideal will hopefully provide greater clarity. Specifically, three necessary and related aspects of practically applicable ideal theorizing are that it is *contextual*, taking into account physical, historical, and sociological limitations, that it is *stable*, and that it is *feasible,* presenting what Rawls refers to as a "realistic utopia."

One line of criticism against ideal theory is that it provides out-of-touch and "context-free political prescriptions."[89] This is certainly true to some extent—a lack of context is what makes ideal theory ideal in the first place. But it must be noted that this lack of context is by no means complete; an applicable vision of ideal institutions must represent, as Mark Jensen puts it, "a picture of the world where our highest aspirations for human society are balanced by our understanding of what humans can actually achieve."[90] Rawls, for example, makes it clear that his theory of justice is limited not just by physical and psychological necessities, but also by cultural and historical constraints. He presents an ideal conception of justice not in some transcendental and timeless sense, but in the context of the modern western constitutional democracies emerging in the wake of various historical events such as the protestant reformation and the liberal attitudes that developed in its wake. The Rawlsian idea of "justice as fairness," he notes, "presents itself not as a conception of justice that is true, but one that can serve as a basis of informed and willing political agreement" which is "securely founded in public political and social

[89] Goodin (1995) p. 56
[90] Mark Jensen, "The Limits of Practical Possibility," *The Journal of Political Philosophy* 17 (2009): 168-184

attitudes."[91] That is, while some inputs are idealized, certain necessary conditions such as the historical context and basic social realities of the society pursuing just arrangements are not.

The idea of *stability* as a necessary feature of an ideal is fairly straightforward and intuitive, but warrants a brief mention. An ideal that is fully achievable for a short time but rapidly devolves into chaos or oppression or creates incentives for individuals to undermine it has very limited desirability. "Other things equal," notes Rawls, "the preferred conception of justice is the most stable one."[92] [93] It seems reasonable to assume, however, that just institutions are to a certain degree self-perpetuating.[94] If individuals feel that a social arrangement is just, there is a natural incentive to maintain it (or, at the very least, not actively oppose it). As Rawls puts it, "if we grow up under a framework of reasonable and just political and social institutions, we shall affirm those institutions when we in our turn come of age, and they will endure over time."[95]

Closely related to this idea of contextual limitation is the idea of *feasibility*. A practically applicable institutional ideal, by taking into account certain necessary constraints, should be realistically achievable. Here we encounter again the distinction explored in the first chapter between inputs and outputs. The contextual limitations just mentioned constrain *inputs*, while concerns over feasibility deal with

[91] John Rawls, "Justice as Fairness: Political not Metaphysical," *Philosophy & Public Affairs* 14 (1985): 223-251
[92] Rawls, *ToJ* p. 498
[93] For another account of stability as a necessary ideal attribute see: G.A. Cohen, *Why Not Socialism?* (Princeton: Princeton University Press, 2009)
[94] The process of just arrangements reinforcing themselves can certainly be analyzed in terms of the social mechanisms of path dependence, though I will not go into it here.
[95] *LoP* p. 7

outputs. As I noted in the explanation of the input-output distinction, the feasibility of outputs is not *necessarily* correlated with the factuality of inputs. In other words, while there are some situations in which the presence of a nonfactual assumption as an input will make the output ideal arrangement infeasible, this is not always the case. In some cases a nonfactual assumption may still provide a perfectly realizable ideal. Given this, a critique of ideal theory based on the existence nonfactual assumptions such as perfect compliance does not hold if the assumption does not jeopardize the feasibility of the output.

This discussion of the relationship between contextual constraints on inputs and the feasibility of outputs leads to an inescapable question: what does it mean to say that an ideal is "feasible" or "realistic"? If NT theory is to attempt to guide institutional development toward the realization of a distant ideal arrangement, how can we evaluate whether the goal can actually be reached? In order to develop answers to these questions I embark in the next section on an exploration of the nature of feasibility.

5. Feasibility

5.1: Considerations of Feasibility

Different ways of evaluating feasibility must be understood as part of a continuous spectrum rather than as a simple and objective feasible-infeasible dichotomy. While a given set of feasibility criteria might provide a clear line which divides the two, the nature of those criteria can vary greatly. Additionally, each side of the feasible-infeasible split contains varying degrees of (in)feasibility. Two alternatives which are both judged to be feasible can still be compared to each other, with one appearing *more* feasible than the other.

On one end of the feasibility evaluation spectrum lies rigidly fact constrained options which take *all* present limitations as insurmountable. At this extreme one might, for example, say that given the present difficulty of electoral reform in the United States, all feasible plans of political action should take the current system as a rigid constraint. Similarly, one might say of Indian politics that the prevalence of corruption means that corruption must be taken as a background condition of any attempt to promote justice. The danger here, clearly, is that such positions strongly favor the status quo and automatically consider plans that involve the abolition or reform of deeply entrenched institutions to be "infeasible."[96] At the other end of the

[96] Incorrectly treating present constraints as immutable can lead to the problem of what Roberto Unger describes as "false necessity," in which decision-making is constrained unnecessarily by treating as fixed that which can be changed. See: *False Necessity* (Cambridge: Cambridge University Press: 1987)

feasibility evaluation spectrum lie positions that do not take present limitations into account at all. Plans for alleviating widespread global poverty through a worldwide revolution of the proletariat might fall into this category. At this extreme, political limitations, such as the conditioned apathy and ignorance of most populations and the political entrenchment of capitalist interests, are ignored. No present factual constraints outside of basic physical and logical necessities are considered to be *necessary* constraints. This is what one might call impractical utopianism. These two evaluations of feasibility can be thought of as *rigid* and *loose* respectively.

Clearly, very few serious plans reside at either end of this spectrum. Any normative theory of justice, be it comparative or ideal, is based on the idea that change is possible. That is, present constraints are not *completely* insurmountable. One can plan for the removal of certain constraints without sacrificing feasibility. However, *some* constraints on feasibility must be accepted if any practical value is to be drawn from an ideal theory. A balance must be struck between rigid adherence to immediate constraints and a looser allowance for constraints to change or dis-appear with time. The question, then, is not about a simple dichotomy between realism and idealism; it is about inquiring into *the extent that present realities can be ignored without sacrificing feasibility*. Where on the spectrum of feasibility must a theory fall to be practical or valuable? And, relatedly, what *types* of factual limitations limit feasibility? I explore now a few important considerations of practical possibility, drawing heavily on the framework provided in Mark Jensen's insightful examination of the subject, before reflecting on their role in NT theory.

Beginning with the most basic and least controversial limitations on feasibility, there are what Jensen describes as *logical* and *nomological* restrictions.[97] The former is the idea that realistic ideals must be logically consistent in a broad sense, e.g. not dependent on propositions like 2+2=5 or the existence of square circles. The nomological restriction is the simple idea that a practical ideal must be physically possible, e.g. not based on the idea that members of the ideal society have perfect knowledge of the universe or the ability to violate the laws of thermodynamics. These two basic limitations, he notes, are "necessary, but not sufficient, for practical possibility."[98] Here, then, is the first set of limitations on ideal theory (and any progressive theory in general). They can be described as *hard constraints*.[99] Practically possible goals *must* take these limitations into account, but practical possibility requires more than just these basic constraints.

In addition to these, Jensen outlines two more limitations on practical possibility. The first—historical necessity—leaves a bit more room for disagreement than the preceding constraints, but the core idea that the world has a fixed history that we can for the most part agree upon is relatively uncontroversial. Any practical ideal must take into account the historical circumstances that shape present conditions and will continue to shape future political trajectories. Historical limitations are, of course, not permanent. They tend to weaken both with time and with gradual social movement away from particular historical understandings.

---

[97] Jensen (2009) pp. 170-171
[98] *Ibid.*
[99] The hard-soft classification is borrowed from Pablo Gilabert & Holly Lawford-Smith, "Political Feasibility: A Conceptual Exploration," *Political Studies* (forthcoming 2012)

The final limitation presented is that of human abilities.[100] This, it seems, is the truly controversial aspect of the feasibility of ideal plans. Can people cooperate peacefully without a sovereign authority? Are economic incentives necessary for technological progress? Can individuals truly extract their reasoning about justice from their present and material circumstances? These are all questions that will fundamentally affect the nature of a feasible ideal, but they are also questions that do not have clear or uncontroversial answers. It is in this fourth area, that of human capability, that an exposition of the nature of the feasible or the practical must proceed.

These questions of human ability, and to a lesser extent historical contingencies, are *soft constraints,* described by Gilabert as "subject to dynamic variation: not everything that is less feasible now (in the comparative sense) need be as infeasible later. Although it is normally difficult to overpower them *now*, it is possible to transform or dissolve them so that they are no longer constraints at some future time."[101] This understanding highlights, I think, two important aspects of the preceding examination of path dependent outcomes.

First, the series of choices leading to an ideal state is constantly changing and developing. Paths toward the ideal are not merely chains of choices within a certain fixed set of constraints; they are chains of choices in which the constraints on each path and at each point changes with time. Two institutional arrangements that are identical at one point in time may even have different possibilities from that point

[100] Jensen (2009)
[101] Gilabert and Lawford-Smith (2012)

onward due to differences in the developmental paths that led each one to that point. This creates much more flexibility in evaluating the feasibility of an ideal and the path toward it. We must examine not only whether there seems to be a possible path to an ideal from where we stand (given present constraints), but also if there is a path that develops in such a way that constraints are dissolved in future.

The second implication I draw from an understanding of dynamic feasibility returns to this idea of NT theory as a solution to the epistemological blindness of attempts at second-best solutions. Recall that in that discussion, like this one, the focus was on removing real world constraints on ideal arrangements. That is, rather than taking all constraints as hard constraints and seeking second-best solutions within these limitations, NT theory attempts to understand progress toward justice as the development of arrangements that will allow for the removal of the constraints themselves. It is no coincidence that the word "constraint" is used to describe both the real world limitations on ideal conceptions *and* the limits of feasible ideas. They are one and the same; but, importantly, they are *soft* constraints.

The central idea to take from the discussion in this subsection is that the feasibility of a fully just institutional ideal is best understood not as an ambiguous question of whether or not it violates present constraints, but instead as a question of (1) whether or not it violates *hard* constraints and (2) how likely it is that the limiting soft constraints can actually be changed.

5.2: Second-Order Abilities and Soft Constraints

Continuing in the vein of Jensen's lucid exposition of the categories of feasibility, I turn now to his analysis of the three types of human ability. These can be classified as *immediate, first order,* and *second order* abilities.[102] Immediate abilities are simply what our hypothetical subject John can do right now. If John has a typewriter in front of him then he has the immediate ability to type out a message. If, however, John has no typewriter but has money to buy one, or a friend he can borrow one from, then we can say that he has a *first order* ability to type out a message. That is, he does not have the immediate ability to do X, but does have the immediate ability to acquire Y (i.e. buy a typewriter or visit his friend) which will in turn allow him to do X (type a message). This ability is first order in that John presently has the internal ability to type a message, but merely lacks the tools. In this sense, the nature of feasibility already clearly extends beyond immediate abilities. Treating immediate abilities as the limit of progressive action would be to accept the status quo to an extreme degree— falling on the extreme fact-constrained end of the feasibility spectrum mentioned earlier. If one cannot get a third party candidate elected to congress immediately, but one *could* with sufficient voter mobilization, then the election of such a candidate, while perhaps difficult, could hardly be considered completely *infeasible*.

[102] Jensen (2009) p. 176

What arises, then, is the question of how many levels of separation separate the feasible from the infeasible. *Second order* abilities, in Jensen's definition, go a step beyond first order abilities in that a basic ability must first be acquired before something becomes even a first order ability.[103] That is, if doing X requires that we first be able to do Y, then X can be said to be a second order ability. If John wants to type out a message in French, then he not only lacks the tools but also the basic ability (knowledge of French) to accomplish the task. He can, however, learn French with time and practice. In this sense he has a second order ability to type a message in French. Note also that John need not actually learn French in order to have the second order ability to do so. To clarify:

-I have the *immediate* ability to do X if I can perform X now.
-I have the *first order* ability to do X if I can do X later with my present skills.
-I have a *second order* ability to do X if I can do X later provided that I am able to do Y first. That is, I do not merely need materials, I need new *abilities* in order to do X.

With this three part classification in mind, I proceed now to bring together the ideas presented so far in this section into a new integrated account of the nature of feasibility constraints. *Hard* constraints, such as physical and logical limitations, can be understood as limiting alternatives to immediate abilities only. Basic logical requirements like the idea that something cannot be, in an absolute sense, both *P*

[103] *Ibid.*

and *not-P* cannot be circumvented by first or second order abilities. That is, there are no situations in which it is possible to be both P and not-P *if* I do Q first. There is no action or ability Q which allows for the violation of logical identities, and thus there are no first or second order abilities to circumvent them; if I do not have the immediate ability to do X in terms of hard constraints, I do not have the ability to do X at all. This limitation of abilities to the immediate is the defining feature of hard constraints.

Soft constraints, on the other hand, are much more flexible; all three types of ability may apply. I may not have the immediate ability to do X, but I might be able to do X with my present skillset if I had certain additional materials (first order) or if I first did or learned how to do Y (second order). It is in the realm of soft constraints that we are faced with a scalar, rather than a binary measure of feasibility. Gilabert notes this shift in feasibility judgments between hard and soft constraints in his designation of two tests of feasibility, which I paraphrase here:[104]

First, the *binary* test states that it is feasible to do X if doing X does not violate any hard constraints. There is no room for degrees of feasibility in this measure; violating even one hard constraint makes a proposition entirely infeasible on the one hand, and on the other any feasible alternative in a binary sense must violate *no* hard constraints (leaving no room for degrees of feasibility).

[104] Gilabert and Lawford-Smith (2012)

Second, the *scalar* test creates an identity between feasibility and probability when comparing alternatives that are feasible in a binary sense, but which still face soft constraints. This test states that it *more* feasible to do X than it is to do Y if, given a set of soft constraints, it is more likely that someone can successfully do X than Y if she tries. This measure of feasibility can exist on a scale and thus allows for comparative judgments (i.e. two alternatives that are feasible in the binary sense can still have different degrees of feasibility)

With this distinction in mind, I move next to the relationship between this account of feasibility and the path dependent aspects of NT theory. There are three important ideas to carry with us: (1) assuming the conditions of binary feasibility have been met, the scalar feasibility of realizing a goal is dependent not only on immediate, but also on first and second order abilities, (2) second order abilities can be understood as forming chains of discreet actions or decisions leading from the present to future realization of states or abilities, and (3) scalar feasibility can be understood as a function of probability.

5.3: Feasibility and Transitional Theory

If the idea of making judgments about a goal that is not immediately achievable based on an evaluation of chains of discrete events leading from the present to the destination sounds familiar, it should. There is a very close connection between

the branching structure of path dependent outcomes and the chains of necessary preconditions present in distant second order abilities. Putting propositions of each type next to each other reveals a remarkable symmetry:

Feasibility: I can do X if I first do Y, and I can do Y if I first do Z. Thus, I have the second order ability to do X if I have the ability now to do Z.

Path Dependence: We can achieve arrangement C (an ideal) if we first achieve arrangement B, and we can achieve arrangement B if we first achieve arrangement A. Thus, choosing A allows for the possibility of achieving C (i.e. does not create a dead end).

Feasibility constraints, understood in this way, provide insight into the nature of evaluation based on future trajectories toward just institutions. On a basic level, transitional ideal guidance takes as its foundational assumption the idea that we have a second order ability to achieve the ideal goal. That is, it can be reached but there are other actions that must be taken before it can become *immediately* achievable. Put differently: it assumes that there exists an unbroken chain of intermediate states between the present and the ideal. This, you'll recall, is the basis of the negative role of NT theory—evaluating choices based on whether or not they allow for the future realization of the ideal. That is, whether or not a given alternative *preserves the second order ability to achieve the ideal.*

The relationship between the progression of social arrangements and the progression of intermediate abilities between a starting point and a goal also highlights the idea, explored in chapter 2, that when using an ideal to evaluate present alternatives mere similarity is not a legitimate metric; a different *type* of relationship must be used to derive guidance from an ideal. Similarly, in evaluating second order abilities the similarity of any intermediate ability to the ideal does not matter; what matters is the preservation of the ability to reach the final action or state. This concept can be made intuitively clear through an example.

Say I want to be able to play one of Chopin's nocturnes on the piano. This is my ideal state, but I have never played a piano or any other musical instrument in my life. Few would disagree that I have the second order ability to play this piece. That is, I could potentially play it given that I first reach a series of intermediate states. I might have to first get a job to make money, then live frugally for months to save up enough to buy a piano, then hire a piano teacher and give up time to practice every day for months before I am finally able to play the piece. Understood in this way as a series of steps to the ideal, it is clear that each individual step does not necessarily make my present condition more similar to the ideal. In fact, steps like getting a job and living frugally might decrease my quality of life without bringing me closer to playing a nocturne in a comparative sense. The benefit of these prudent actions is instead *transitional*. Getting a job may be unpleasant, but it has transitional value in that it ensures that my second order ability to play a nocturne is more likely to be realized.

This analogy can be examined through both *linear* transitional and comparative lenses as well. From a linear transitional standpoint (in which the best alternative is the one that most resembles the ideal) the best option might be to buy a 25 dollar plastic keyboard on which I can learn to play 'chopsticks' in ten minutes. In terms of similarity to the ideal of playing Chopin, this certainly seems to lie closer than filling out and submitting a job application. Yet in a transitional sense it is not particularly useful at all.

Alternatively, a strictly comparative evaluation might consider the distant second order ability to play a nocturne too far away, and focus instead on remedying other problems which have immediate, rather than second order, solutions. For example, fixing a leak in the roof of my house might be comparatively more beneficial than getting an unpleasant job with the distant goal of playing Chopin, even if Chopin would ultimately bring me greater happiness than fixing a small leak in the roof. Once the leak was fixed, I might move then to mending a hole in my trousers or getting a new pair of shoes or some other tangibly beneficial task. In short, by focusing strictly on comparative evaluations I might move from one immediately accessible solution to the next, always choosing the alternative that has the greatest short-term benefit and never taking the necessary sacrificial steps toward a higher ideal.

Moving now from my ideal of playing Chopin to institutional ideals of justice in a somewhat Platonic analogy of the individual and society, it appears that both purely comparative and linear transitional methods of evaluating alternatives are

biased toward immediate or first order abilities with the result that they do not take into the account the possibility of dead ends in the pursuit of second order abilities (which institutional ideals unavoidably are in our present world). The account of NT theory laid out in chapter 2, understood in the present context, avoids the second-best issues of linear comparisons to the ideal by using a *metric of feasibility*. Ignoring the short-term desirability of alternatives (above a certain moral threshold), it pursues paths of potentially austere social arrangements for the sake of preserving the second order ability to achieve fully just arrangements.

The idea of avoiding dead ends and favoring alternatives that leave open the possibility of the eventual realization of ideal institutions can be thought of, as I've said, as the preservation of second order abilities. Referring back to the visualization of dead ends in the PD tree in §2 of this chapter (fig.2 p.84), the various chronological stages (A, B, and C) can be thought of as orders of separation from the ideal. From the starting position we had an immediate ability to do any of the "A" options, a first (or second) order ability to do any of the "B" options, and a second order ability to do any of the "C" options.

The choice of the most immediate desirable option, (A1), you'll recall placed certain limitations on future abilities. From (A1) we were able to choose either (B2) or (B4), the second and fourth best options overall in the second round of alternatives. Thus, after choosing (A1) from the starting position, we lost the ability to reach (B1), (B3), and (B5). As a result, then, we lost the ability to achieve all of the arrangements that resulted only from those three foregone possibilities. The core

idea to take away from this process is that every policy choice not only foregoes other immediate possibilities, but also disallows the realization of further first and second order abilities.

The connection between feasibility and probability mentioned near the end of the previous subsection warrants further examination insofar as it contributes to the NT evaluation of alternatives. The broad and frequently negative role of NT theory is generally to make categorical claims as to whether or not a policy alternative preserves or does not preserve the second order ability to achieve a given ideal. I refer to it as a "negative role" in the sense that, due to epistemological limitations, such evaluations will more frequently be claims that a given alternative *does not* preserve a path to the ideal than that a given alternative definitely *does* allow for the realization of the ideal. Both claims, of course, will necessarily fall far short of certainty. But the former, it seems, can be supported more strongly than the latter.

For example, we can say with more certainty that a hereditary aristocracy will *not* allow for egalitarian distributive justice than we can that an inclusive democracy *will* allow for such an ideal. This negative role of NT theory demonstrates well the nature of probability as a scalar measure of feasibility. Both the hereditary dictatorship and inclusive democracy might preserve the second order ability for an egalitarian distribution of resources, and as such NT theory might not rule either one out as a dead end. The higher probability of achieving the ideal in a democracy, however, certainly should be taken into account. In this sense, the option of

democratic institutions is more desirable not merely in a comparative, but also in a transitional sense.

I conclude the present examination of feasibility by returning, as I did in chapter 2, to an evaluation of the claim that ordinary moral ideals can function just as well as institutional ideals in a nonlinear transitional theory. Evaluating second order abilities to achieve a given moral ideal, one might argue, can provide guidance based on general moral principles without an appeal to the "baggage" of ideal theory. However, in this case too we find that the specificity of ideal theory, which some interpret as a problematically rigid, is precisely what allows for the transitional metric to be used. Such a metric is, as I have argued, a necessary step in overcoming the apparent blindness created by evaluating second-best solutions in constrained circumstances.

Broad moral claims are certainly necessary for evaluating immediate alternatives to ensure that they remain above a certain threshold of moral acceptability, or for making comparative judgments between alternatives that are equally acceptable by NT standards. However, when evaluating second order abilities (i.e. the existence of an unbroken progression of social arrangements leading to the ideal), it is only by deriving institutional ideals from moral ones that a clear binary distinction can be drawn between conditions being "met" or "unmet." This clarity is essential if the necessarily fuzzy predictions of long term second order abilities can be evaluated with even the smallest amount of clarity.

Reflecting back as the present line of argument draws to a close, there are a few central points to be drawn from this chapter's discussion. First, the existence of path dependencies in the progression of institutional arrangements necessitates consideration of the future limitations created by present choices. As a result, a nonlinear transitional evaluation of justice (which I argue is valuable in light of the seeming impossibility of second-best limitations) must consider the transitional realization of just outcomes as the procession of a series of branching alternatives, with some choices preserving unbroken chains of development toward the ideal. An evaluative framework of this kind will sometimes recommend against the most immediately desirable alternative if it appears that such a choice will lead to a *dead end*, a state from which there are no foreseeable paths to the ideal.

Second, the similarity between discussions of feasibility and path dependence allows for a deeper understanding of both. A dead end state can, in the language of feasibility, be understood as a state in which the second order ability to achieve a given ideal is no longer likely to exist. Chains of second order feasibility, it appears, are strikingly similar to a nonlinear transitional method of evaluation in that each individual step is necessary or desirable not because it resembles the end goal, but because it preserves the ability to move through a progression of actions or states leading ultimately to that goal. Thus NT theory, which became necessary due to the invalidity of comparing constrained arrangements based on similarity to the ideal, appears also to be the basis of evaluation for both path dependent progressions and measures of feasibility.

We are inescapably limited in our ability to judge the "best" policy choice in a given situation. But if the preceding argument is sound, then NT theory provides the best framework for using a realistic but distant ideal in the prudent evaluation of present and immediately achievable social arrangements.

6. Utilitarian and Authoritarian Critiques

I take a moment now to examine two significant objections to the potentially austere and sacrificial nature of policy decisions informed by NT considerations. These strands of criticism against NT theory which warrant special attention can be described as the *utilitarian* and *authoritarian* critiques. Beginning with the utilitarian critique, it should be noted that implicit in such a critique is an assumption about the (un)desirability of utilitarianism. That is, to critique a theory of applied ideal justice on the grounds that it is utilitarian implies that utilitarianism itself is unjust. The approach I have been presenting, however, is not based on any one particular conception of justice. It is, in fact, perfectly compatible with a utilitarian view of justice. In such a situation utilitarian tendencies would clearly not be considered a flaw.

But, if one accepts the view that utilitarian tendencies are in fact problematic, the apparent willingness of NT theorizing to make present sacrifices for future gains

toward justice might be considered a serious flaw. Even if the content of the ideal is in no way utilitarian, it might be argued that this method of progress toward it inescapably is. The value placed on the realization of justice in the future implicitly puts the quality of life of the many (i.e. future generations) over the few (those currently alive and making decisions). To repeat my earlier example: an industrial worker in the 19th century would find little consolation in the fact that his toil and suffering will make possible the development of an economic order that allows for improvements in social justice in the next century. Similarly, a slave in a utilitarian system would find little consolation in the fact that the loss of utility in his life is offset by the gains of a hundred others that benefit from his labor. If individuals are to be treated as ends in themselves in the Kantian tradition, rather than as means toward justice, what justification is there for not choosing the most comparatively desirable alternatives for those living in the present?

This utilitarian concern meets some resistance from a frequently recurring topic in contemporary discussions of justice: *intergenerational justice*. What responsibilities do those currently alive have to future generations in terms of justice? Increasing the rate of fossil fuel consumption might yield immediate energy benefits, lowering prices and allowing for a higher quality of life for the next few decades. However, if such use not only leaves no fossil fuels for future generations but also causes irreversible climatic and atmospheric changes, it can hardly be described as intergenerationally just. Recommending that present generations forego some of the potential benefits of increased fossil fuel usage in order to allow for a

higher quality of life for future generations is not utilitarianism. It is a basic consideration of intergenerational fairness. In a similar vein, recommending that those presently alive forego the most immediately desirable institutional arrangements if they would limit or degrade the possible choices of future generations is not strictly utilitarian.

Looking at a nearly opposite situation, a strictly libertarian conception of justice might not accept the necessity of any kind of present sacrifice in the name of future benefits. In such a framework maximizing present gains in justice would be an acceptable pursuit under most definitions of basic Lockean natural rights. An argument might be made for intergenerational rights within a libertarian framework, but this is not the place for a such a discussion. The inverse of this is might be a hypothetical genuinely utilitarian NT theory that does not consider present conditions at all in evaluating paths to ideal justice. Such an approach would place no constraints on the moral acceptability of policies, and might recommend the submission of present generations to extreme oppression in the name of future realization of complete justice.

Lying between these extremes of strong libertarianism on the one hand and utilitarianism on the other, NT theorizing may advocate foregoing maximal gains in justice in the present not out of a callous disregard for those presently alive, but out of a necessary regard for those yet to be born. It is restricted on one side from picking the most desirable alternatives without considering future limitations, and

on the other from picking immediately morally unacceptable alternatives even if they facilitate the future development of just arrangements.

Moving now to the *authoritarian* critique, there may be valid concerns about the possibility of politically exploiting an NT approach. In fact, it may well be argued that NT theory has *already* been used to justify massively oppressive actions. The years following the Russian Revolution provide the example *par excellence* of just such oppression in the name of transitional benefits. The 1917 shift toward what Lenin called "proletarian democracy" meant the violent oppression of a significant fraction of the population. As he notes in his writings, "violence exerted in the name of interests and rights of the majority of the population…tramples on 'rights' of exploiters—of the bourgeoisie."[105] Yet clearly the blatant oppression of large segments of the population was not some sadistic end in itself; the future social ideal of the incredibly violent dictatorship of the proletariat was the eventual development of a "society without classes, without a state, and consequently without violence."[106]

Compared to this goal, the rise of a soviet dictatorship based on a powerful state and widespread violence certainly seems to be a step away from the ideal in terms of similarity. In terms of desirability as well, especially as time went on and the single-party state grew more and more invasive and heavy-handed, soviet style communism would not have ranked very highly on the preference rankings of many Russians. The somewhat gentler "bourgeois democracy" might in this case appear to

[105] Lenin (Russian ed.), Vol. XXX pp. 260-261. Quoted in: Andrei Vyshinsky, *The Law of the Soviet State*, trans. Hugh Babb. (New York: Macmillan, 1961)
[106] Vyshinsky (1961)

be an improvement in both of these aspects. From Lenin's deeply transitional perspective, however, bourgeois democracy was considered a developmental dead end which, as long as it existed, would prevent the future development of true communism. In this way transitional development was used to justify mass misery and oppression.

The mode of thinking present in NT theorizing can certainly be dangerous when combined with an executive body that has too few moral limitations and too much confidence in potential outcomes. The theory is, in its least flattering form, a justification of injustice and oppression, of invasion and colonialism, of means justified by ends. This use as a tool for justifying oppression, however, in no way discounts the potential benefits of ideal guidance. Claims of this type should, of course, be subjected to careful scrutiny and bound by certain moral limitations. In the following section I explore some of these important constraints on NT theory.

7. The Limits of Nonlinear Transitionalism

With the basic methodology and potential dangers laid out, I end this chapter with a discussion of the necessary limitations of NT theorizing. Adding to the general guidelines for ideal theory laid out at the end of the first chapter, I hope to focus now not on expanding the realm of applicability of ideal theory, but on

assessing its constraints and frontiers. I divide these limitations into four categories, to be addressed in order: the *moral*, the *political*, the *practical*, and the *epistemological*. Although entire books could be (and have been) written about each of these, the limited nature of the present project requires that I settle for a brief sketch of the first two limitations. The latter two, on the other hand, correspond to the discussions present in the third and second chapters, respectively, of this essay.

The fact that nonlinear transitional theory allows for the selection of policies that may require short term sacrifices in order to preserve the second order ability to eventually achieve the ideal state requires of us that we consider how extreme such sacrifices can be. In this sense it becomes clear that while I argue that moral principles alone do not provide the most effective approach to remedying injustice, they still have an essential place in evaluating nonideal alternatives. When nonlinear transitionalism favors more extreme immediate sacrifices, moral limitations must push back and limit how far such sacrifices can go. This is the essentially the first part of Rawls's claim that nonideal theorizing must look for courses of action that are "morally permissible and politically possible as well as likely to be effective."[107] This applies not only to blatantly immoral means (such as murdering political dissenters), but also, as Simmons notes, to institutional "rug pulling" in which "people base life plans or important activities on the reasonable expectation that the rules will remain unchanged…and then have the rug pulled out from beneath them by sudden institutional change."[108] In responding to moral claims like these, NT

[107] *LoP* p. 89
[108] Simmons (2010) pp. 20-21

theorizing is constrained in its pace and methods; consideration for future possibilities must be tempered by immediate moral claims.

The second limitation on the application of NT theory is *political*. One type of necessary constraint in this area is the Rawlsian idea of reasonable pluralism, which consists of the claim that under democratic institutions "a diversity of conflicting and irreconcilable yet reasonable comprehensive doctrines [i.e. belief systems] will come about and persist."[109] The acceptance of rational pluralism may put limits on the pursuit of institutional ideals by providing multiple reasonable conceptions of what those ideal institutions should be. As a result, it might be reasonable to constrict nonlinear transitional modes of development by adding a further condition that decisions should try to preserve feasible paths to *multiple* ideal states. A guideline such as this would obviously have limits when a choice had to be made between mutually exclusive paths. However, while these plural ideal arrangements remained remote one might reasonably require NT theorizing to favor policy paths that sacrificed speed towards one particular ideal for the sake of leaving paths to others open.

The third limiting factor for NT theory consists of the *practical* limitations discussed in this chapter. These are essentially the requirements of feasibility discussed in §5. On the basic and uncontroversial level there are logical and physical requirements that an ideal must conform to. Beyond these there are historical necessities that, while perhaps more contentious in content must undoubtedly be

[109] John Rawls, *Justice as Fairness: A Restatement,* ed. Erin Kelly (Cambridge, Massachusetts: Harvard University Press, 2001)

taken into account. Finally, there are the limitations of human ability, in particular the extent to which second order abilities can reasonably be considered feasible.

Finally, the *epistemological* limitations of ideal theory emerge from the discussion of second best solutions in chapter 2. The nature of second best alternatives heavily limits the possibility of knowing whether actions taken are beneficial in an overall sense—whether they move us closer to second-best optima in constrained conditions. On the one hand, given an interdependent set of ideal conditions in which one or more are constrained, similarity to the ideal fails to serve as an adequate measure of progress. The negative corollary of this, however, is that under constrained conditions with no direct comparison to an ideal possible, we cannot know whether any particular change in any area is beneficial on the whole or over time.

I have presented nonlinear transitional theory as an alternative to this apparent impossibility of informed judgment, but it is important to keep in mind that it still faces serious epistemological challenges. First, NT judgments are based on predicting path dependent outcomes of which there can be tendencies and trends, but no guarantees. Second, NT judgments do not pick the single best alternative in a given set, but rather attempt to identify those alternatives that are consistent with political trajectories that eventually achieve ideal justice. Finally, there will almost certainly be many times when a reasonable judgment about future possibilities is not possible with an acceptable level of confidence. In such cases we may simply be

forced to rely on basic comparative gains, crossing our fingers and hoping for the best.

Chapter IV
Conclusion

> *If then our acting well were to consist in this, in our grasping and acting in accord with the great distances and avoiding the small distances and not acting in accord with them, what means of saving our life would have come to sight? The art of measuring or the power of appearances?*
>
> Socrates, *Protagoras*[110]

1. The Value of Ideal Theory

This essay began with a specific task: to present an account of the practical value of ideal theory. Putting aside much stronger claims as to necessity or sufficiency of ideal guidance for making comparative judgments, I have sought to demonstrate instead that it is a *useful* tool in the pursuit of justice. Against recent arguments that ideal theory is an impractical and purely academic distraction, I argue that there are methods by which ideal theory can be employed not to *replace*, but to *improve* fully fact-constrained and contextual comparative judgments.

The first portion of this argument explored the limitations inherent in trying to discern what the second-best alternative is when one or more conditions of an optimal first-best arrangement cannot presently be met. The broad interpretation of

[110] Plato, *Protagoras,* trans. Robert Bartlett (Ithaca, NY: Cornell University Press, 2004) 356d

the general theory of second best placed serious limitations on the possibility of knowing how to rank a set of available alternatives against each other. On the one hand, when a constraint is introduced on even one of a set of interdependent optimal conditions, similarity to ideal conditions is not a legitimate method of evaluation options. This rules out any *linear* ranking of alternatives, including linear transitional theory. However, the idea that ideal conditions may no longer be desirable in constrained circumstances also puts serious limits on our ability to evaluate whether a given change in a single condition is beneficial at all. This frequently overlooked negative corollary severely hampers efforts to make purely comparative and partial evaluations of social arrangements.

Confronted with this apparent impasse in which neither ideal guidance nor piecemeal comparative strategies seemed capable of making any solid claims whatsoever, I presented *nonlinear transitional theory* as a method of ideal guidance that avoided to some extent the crippling uncertainty of second-best arrangements. By evaluating present options based not on their immediate desirability or similarity to an ideal, but on the degree to which they allowed for the eventual and complete realization of that ideal, NT theory provided a way to make judgments with reference to an ideal without having to rely on impossible measures of similarity.

With this alternative method of ideal guidance in hand, I turned in the third chapter to a discussion of what concrete considerations an NT approach might make in evaluating transitional desirability. The first of these were the mechanisms of path dependent outcomes, through which present choices affect the range of

possible options far into the future. For example, some policies might be heavily self-reinforcing such that once chosen, it becomes very difficult in the future to choose something else. Others might have negative externalities in seemingly unrelated parts of the institutional structure such that choosing a specific course of action in one area might close off the possibility of pursuing another somewhere else. In both of these cases and others, the primary concern of NT theorizing is the recognition of choices that might potentially close off paths to the realization of fully just institutions in the future. These are what I refer to as dead ends. Importantly, an effort to avoid dead ends may in some cases necessitate the pursuit of policies that either appear to move a society away from similarity to ideal conditions or that are immediately undesirable in terms of preference rankings, or both.

The argument turned next to the idea of feasibility; specifically, the idea that feasible goals could be understood in terms of immediate, first-, and second-order abilities. Although compatibility with *hard* constraints such as physical and logical necessities required that the feasible and the immediately achievable be one and the same, I demonstrated that compatibility with mutable *soft* constraints required only that we have at least a second-order ability to achieve a goal in order for it to justifiably be considered feasible. The nature of progress toward such second-order abilities was then compared with the branching decisions of path dependence analyses and found to be fundamentally similar. Both of these processes consisted of a series of discrete actions or decisions that led from the present to the realization of a specific ability or arrangement that was indirectly possible at the start (i.e. a

second order ability) but that could become immediately achievable through a specific series of intermediate events.

The recognition, preservation, and pursuit of this type of path—the type that allows second order abilities to eventually become immediately achievable—is the essential task of nonlinear transitional theory. This method of ideal guidance is quite compatible with our basic individual intuitions about the desirability of looking toward future possibilities and making short term sacrifices for long term gains (which will be discussed in the next section). And, in fact, one can find examples of nonlinear transitional reasoning and justification throughout history, from the Bolsheviks to the European Central Bank. But despite this compatibility with many individual and political intuitions, a rigorous account of the necessary theoretical foundations and nonideal considerations of this type of ideal guidance has until now not been attempted.

Each step in the trajectory of this argument has, hopefully, been demonstrated clearly in the preceding chapters and led to an unambiguous conclusion: applying ideal theory to present judgments of the desirability of available social arrangements through the proper application of *nonlinear transitional* methods and considerations is undoubtedly beneficial. NT theory does not, and should not, provide a complete and monistic framework for using institutional ideals alone to judge all available alternatives; immediate moral and political considerations must also be taken into account. But to ignore NT considerations in the pursuit of social justice is to forego a valuable tool in the long-term development of that goal.

Putting aside the specific content of any particular conception of justice, I take this conclusion to be an extension of the basically Rawlsian idea of the place and applicability of ideal theory as a tool for bringing about a fully just basic structure in our presently constrained world. It addresses, at least partly, Rawls's deliberate vagueness as to the exact nature of ideal-guidance based evaluations of immediate choices. "The problems of partial compliance [i.e. nonideal] theory," he notes near the beginning of the *ToJ*, "are the pressing and urgent matters. These are the things we are faced with in everyday life." However, "the reason for beginning with ideal theory is that it provides, I believe, the only basis for the *systematic grasp* of these more pressing problems."[111] But despite this intuitive belief in the possibility of practical ideal guidance, Rawls and those that have followed after him never articulated the precise nature of this "systematic grasp". But if the arguments presented in the preceding chapters are sound, my account of nonlinear transitional theory is the method of practical ideal guidance through which such a systematic grasp is possible.

With the main body of my central thesis concluded, I would like to take the rest of this final chapter to briefly examine the connection between NT theory and our intuitions regarding prudent action. This account is not essential to the main argument, but it will hopefully highlight the reasons why the method of ideal guidance presented in this essay should not be understood as an inaccessibly academic

[111] *ToJ* p. 9. Emphasis mine.

and analytic framework suitable only for books and journals. Nonlinear transitional theory certainly has a robust theoretical grounding that can be extended far beyond the concepts in this essay; however, *it is an approach that is fundamentally rooted in our natural intuitions and ideas about what it means to act with prudence and foresight in the pursuit of future gains.*

It is not a theory that political actors would have to study carefully in order to grasp intuitively. The intuition is already there, and can be captured in two words: *delayed gratification*. The heart of NT theory lies in our basic rational capacity for planning ahead—our willingness to forego something we want today in order to gain something we want even more tomorrow. As a result, the cultivation of NT modes of thinking in society does not require the futile task of undermining and replacing basic human intuitions. Instead, it takes these intuitions as a starting point and *extends* them, drawing on common-sense views about what it means to act well on an individual level and applying them to political action on the grandest scale.

2. NT Theory as Social Prudence

One way of understanding nonlinear transitional theorizing is as a theory of what can be thought of as *social prudence.* It is, in this sense, a theory that provides a framework for long term decision-making that seeks to balance present desirability

and long term benefit. Adam Smith, in his *Theory of Moral Sentiments,* provides a useful framework for an understanding of this sort with his idea of the *impartial spectator*. The impartial spectator, for Smith, plays the role of our internal conscience; he is "the inhabitant of the breast, the man within, the great judge and arbiter of our conduct."[112] By imagining ourselves in the position of an impartial spectator, who does not share in our individual sentiments, we are able to overpower (at least partly) our natural bias toward individual passions and desires. We are, as a result, able to see "the propriety of resigning the greatest interests of our own, for the yet greater interests of others, and the deformity of doing the smallest injury to another, in order to obtain the greatest benefit to ourselves."[113]

In the position of the impartial spectator we are able to observe our actions and choices from a more general standpoint—to see ourselves through the eyes of others and to judge our actions accordingly. Importantly, this account of the impartial spectator is not Smith's attempt to implore readers to consider outside perspectives; it is an empirical account of how we actually *do* evaluate our actions by considering how someone else who did not share in our immediate feelings might judge us if they were watching. In this way Smith does not attempt to lay out a moral system that runs contrary to natural human inclinations; rather, he observes those natural tendencies and attempts to systematize the moral sentiments that we already have. In this way, Smith's project is in at least one sense a model for my own

[112] Adam Smith, *The Theory of Moral Sentiments* (Indianapolis: Liberty Fund, 1982)(Hereafter *TMS*) III.3.4
[113] *TMS* III.3.4

attempt to develop a theory of ideal guidance firmly grounded in natural intuitions about acting with an eye to the future.

Amartya Sen has recently used Smith's idea of the impartial spectator to propose a method of avoiding national parochialism in comparative judgments of justice.[114] There is certainly room for criticism of this attempt to free the content of moral judgments from the apparently inescapable localism that naturally results from the fact that every individual develops within, and is necessarily conditioned by, particular social environments. Such a discussion, however, would revolve around the nature and limitations of the moral content of a theory of justice, which is not the focus of this essay.[115] I would instead like to use the idea of a Smithian conscience to examine the ways in which we avoid not spatial or cultural, but *temporal* parochialism. Importantly, the problem of social and moral localism is to some extent avoided by the fact that such temporal considerations do not attempt to step outside existing moral judgments. They do not look beyond the present conception of the good; they instead attempt to look beyond present ideas of how best to achieve that good in the future. Avoiding temporal bias, then, is more like the "art of measuring" mentioned by Socrates at the beginning of this chapter, while the possibility of avoiding moral or cultural bias creates a much more serious challenge. The impartial spectator, in the context of temporal judgments, provides a useful framework for understanding the intuitive appeal of a nonlinear transitional

[114] Sen (2009)
[115] For an in-depth account of the issue of moral and sympathetic parochialism in Smith's thought see: Fonna Forman-Barzilai, *Adam Smith and the Circles of Sympathy* (Cambridge: Cambridge University Press, 2009)

framework, first as a question of *intergenerational justice* and second as a question of *social prudence*.

In evaluating alternative social arrangements in terms of justice, there is frequently an inclination, as in most things, to favor immediate and individual gratification. This naturally runs counter to our intuitions about the desirability of considering long-term consequences. It is in response to our myopic inclinations that the idea of intergenerational justice emerges. The view that we should evaluate arrangements not merely in terms of present desirability, but also with an eye to the effects on future generations depends on our ability to judge from a temporally neutral standpoint that does not favor the present over the future (or, at least, takes the future into account). The interests of future generations cannot, from a standpoint focused entirely on present desirability, "be put into the balance with our own, can never restrain us from doing whatever may tend to promote our own, how ruinous soever to [others]."[116] Such restrictions on the pursuit of present desirability are an essential part of nonlinear transitional theorizing, which looks not merely at comparative gains in justice, but also at improving prospects for the future realization of just social arrangements that may extend generations into the future. NT theory is, in this understanding, the institutional manifestation of intergenerational justice.

This intergenerational way of thinking captures some of the intuitions behind *why* transitional justice is important. But in developing also a method for

[116] *TMS* III.3.3

considering *how* transitional justice is to be realized, NT theory can be understood also as a theory of what I call social prudence. Prudence, on both the individual and social level, can be though of essentially as the ability to *overcome temporal bias*[117] in judging what actions to take in the pursuit of a given end—to give equal weight to the far and the near despite the fact that the latter always appears greater. One can trace the roots of this idea of prudence as far back as Plato's *Protagoras* in which Socrates observes, in his discussion of the best way to pursue pleasure in life, that pleasures of identical magnitudes appear greater when they are near and further when far away.[118] Given this, he argues that "acting well" consists in the ability to accurately measure the consequences of one's choices "in accord with the great distances."[119] The opposite of this would consist, then, in making intuitive judgments based on immediate circumstances without an eye to a long term goal. Such an unwillingness to consider comprehensive and long-term possibilities would cause us, as he says, "to wander about and change our minds back and forth many times about the same things and go back on our decision when it comes to both our actions and our choosing things that are great and small."[120] [121]Prudence, in the Socratic view, consists in not only considering long term states but in giving them *equal* weight—in overcoming the natural distortion of present and future states that results from our particular temporal perspectives.

---

[117] Etymologically, *prudence* comes to us from Latin *prudentia*, itself a contraction of *providentia,* lit. *seeing ahead.*
[118] Plato, *Protagoras* 356c
[119] *Ibid.* 356d
[120] *Ibid.*
[121] One can hear echoes of this in Rawls's description of intuitionism as "[consisting] of a plurality of first principles which may conflict to give contrary directives in particular types of cases." *ToJ* p. 34

This basic idea can be found also in Smith's discussion of prudence and the way in which temporal bias is overcome. Echoing Socrates, he uses physical distance as an illustrative example of the way in which our individual perspective distorts our estimation of that which is large but distant, then describes how it is overcome:

> I can form a just comparison between those great objects and the little objects around me, in no other way, than by transporting myself, at least in fancy, to a different station, from whence I can survey both at nearly equal distances, and thereby form some judgment of their real proportions.[122]

This "different station" from which the biases of visual perspective can be overcome in a physical sense manifests itself on a more personal level as the "man within the breast," the impartial spectator through which we attempt to judge without the distortions of individual sentiment. Attempting to judge from a more impartial standpoint has, on the one hand, *interpersonal* effects which change the way we act on our own interests, desires, and feelings by weighing them against those of others. On an *individual* level, however, when acting prudently we tend also to try to take an impartial position in order to accurately judge present and future pleasures or sentiments. This temporal impartiality is an important feature of Smithian prudence. The prudent man, for Smith, "is always both supported and

[122] *TMS* III.3.2

rewarded by the entire approbation of the impartial spectator" who, because he does not partake in any particular moment's sentiments or appetites, sees the present and the future "nearly at the same distance, and is affected by them very nearly in the same manner."[123]

Given a choice between something good now and something great later, the prudent man will, having compared them without the bias of our natural desire for immediate gratification, choose the latter. Analogously, given a choice between moderate gains in justice now and the possibility of a fully just basic structure in the future, NT theory will, having evaluated them without a bias toward immediate comparative gains, choose the latter.

Thus while individual prudence can be thought of as the temporally unbiased pursuit of individual ends such as pleasure, wealth, rank, happiness, etc., the social prudence of NT theory attempts to realize this broader perspective in the development of *social* ends. The end presently under consideration is justice, but this mode of thinking need not be limited to such a pursuit. I present NT theory as a way of promoting long-term progress toward *any* comprehensive social-institutional ideal, including an ideal that has nothing to do with justice. It is a strategy, not an end.[124]

By rooting this theory in basic human intuitions about prudent action and delayed gratification, I hope to have brought together a system of understanding that grows naturally in the mind and needs only description and rigorous analysis—not counterintuitive and academic persuasion—in order to flourish. In this way I

[123] *TMS* VI.i.11

[124] Or, if it truly is analogous to prudence, one might be so bold as to call it a virtue of political action.

align myself, to some extent, with what Samuel Fleischacker describes as "a central commitment, running through all [of Adam Smith's] work, to vindicating ordinary people's judgments, and fending off attempts by philosophers and policy-makers to replace those judgments with the supposedly better 'systems' invented by intellectuals."[125]

These theoretical foundations of transitional ideal guidance are, in important ways, attempts to capture and systematize intuitions that are often already present in everyday political and domestic considerations. Returning to my earlier examples, it seems natural that I would consider courses of action that did not include piano lessons, or that focused only on an endless procession of small and immediate concerns, to be dead ends in the pursuit of my goal of playing Chopin. It seems perfectly reasonable that I would sacrifice some of my leisure time now in order to work enough to afford piano lessons in the future. It seems rather intuitive that if I can't have both milk and cookies, I might not want either alone—or that my not being able to immediately do something doesn't mean that it is an infeasible goal. All of these examples demonstrate the natural accessibility of the central concepts of NT ideal guidance.

The *political* realization of this way of thinking, however, requires something more than individual intuition: it requires mutual understanding and public reason, as well as collective action. It requires also that the judgments we make internally and automatically every day about how best to weigh present and future benefits be

[125] Samuel Fleischacker, *On Adam Smith's Wealth of Nations* (Princeton: Princeton University Press, 2004)

verbalized, analyzed, and debated. Transitional methods of reasoning underlie many political decisions, but usually remain buried under rhetoric. In explicating the nature and importance of these judgments in the realm of politics and policy I hope to have presented a way of bringing these implicit considerations of future ideals and the present sacrifices they entail to the surface—and to have outlined the beginnings of path toward revitalizing the decaying relationship between political action and political vision.

References

Baumol, William J. 1965. "Informed Judgment, Rigorous Theory, and Public Policy," *Southern Economic Journal* 2: pp. 137-145.

Cohen, G.A. 2009. *Why Not Socialism?* Princeton: Princeton University Press.

Farrelly, Colin. 2007. "Justice in Ideal Theory: A Refutation," *Political Studies* 55: 844-864.

Fleischacker, Samuel. 2004. *On Adam Smith's Wealth of Nations.* Princeton: Princeton University Press.

Forman-Barzilai, Fonna. 2009. *Adam Smith and the Circles of Sympathy.* Cambridge: Cambridge University Press.

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

Goodin, Robert. 1995. "Political Ideals and Political Practice," *British Journal of Political Science* 25: 37-56.

Jensen, Mark. 2009. "The Limits of Practical Possibility," *The Journal of Political Philosophy* 17: 168-184.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor. Cambridge: Cambridge University Press, 1998 [1785].

Lipsey, R.G. and Kelvin Lancaster. 1956. "The General Theory of Second Best," *The Review of Economic Studies* 24: 11-33.

Machiavelli, Niccolo. *The Prince*, trans. Russell Price. Cambridge: Cambridge University Press, 2008 [1532].

Mills, Charles W. 2005. "'Ideal Theory' as Ideology," *Hypatia* 20: 165-184.

Mishan, E.J. 1962."Second Thoughts on Second Best," *Oxford Economic Papers* 14: 205-217.

Morrison, Clarence. 1965. "The Nature of Second Best," *Southern Economic Journal* 32: 49-52.

Nagel, Thomas. 2005. "The Problem of Global Justice," *Philosophy & Public Affairs* 33: 113-147.

O'Neill, Onora. 1996. *Towards Justice and Virtue*. Cambridge: Cambridge University Press.

Page, Scott E. 2006. "Path Dependence," *Quarterly Journal of Political Science* 1: 87-115.

Plato. *Protagoras,* trans. Robert Bartlett. Ithaca, New York: Cornell University Press, 2004.

Plato. *Republic*, trans. G.M.A. Grube. Indianapolis: Hackett Publishing Company, 1992.

Rawls, John. 1971. *A Theory of Justice* (*ToJ*). Cambridge, Massachusetts: Harvard University Press.

Rawls, John. 1985. "Justice as Fairness: Political not Metaphysical," *Philosophy & Public Affairs* 14: 223-251.

Rawls, John. 1999. *The Law of Peoples* (*LoP*). Cambridge, Massachusetts: Harvard University Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement,* ed. Erin Kelly. Cambridge, Massachusetts: Harvard University Press.

Rousseau, Jean-Jacques. *The Social Contract*. In *The Basic Political Writings*, trans. Donald Cress. Indianapolis: Hackett Publishing Company, 1987 [1762] pp. 141-227.

Schmitt, Carl. *Political Theology,* trans. George Schwab. Chicago: University of Chicago Press, 2005 [1922].

Sen, Amartya. 2004. *Rationality and Freedom.* Cambridge, Massachusetts: Harvard University Press.

Sen, Amartya. 2006. "What Do We Want from a Theory of Justice?" *The Journal of Philosophy* 103: 215-238.

Sen, Amartya. 2009. *The Idea of Justice.* Cambridge, Massachusetts: Harvard University Press.

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

Smith, Adam. *The Theory of Moral Sentiments* (*TMS*), ed. D.D. Raphael and A.L. Macfie. Indianapolis: Liberty Fund, 1982 [1790].

Stemplowska, Zofia. 2008. "What's Ideal About Ideal Theory?" *Social Theory and Practice* 34: 319-340.

Swift, Adam. 2008. “The Value of Philosophy in Nonideal Circumstances,” *Social Theory and Practice* 34: 363-387.

Unger, Roberto. 1987. *False Necessity*. Cambridge: Cambridge University Press.

Vyshinsky, Andrei. 1961. *The Law of the Soviet State*. trans. Hugh Babb. New York: Macmillan.

Wiens, David. 2011. “Prescribing Institutions Without Ideal Theory,” *The Journal of Political Philosophy* 20: 45-70.