# A PHILOSOPHER OF SCIENCE LOOKS AT IDEALIZATION IN POLITICAL THEORY*

By Jenann Ismael

Abstract: *Rawls ignited a debate in political theory when he introduced a division between the ideal and nonideal parts of a theory of justice. In the ideal part of the theory, one presents a positive conception of justice in a setting that assumes perfect compliance with the rules of justice. In the nonideal part, one addresses the question of what happens under departures from compliance. Critics of Rawls have attacked his focus on ideal theory as a form of utopianism, and have argued that political theory should be focused instead on providing solutions to the manifest injustices of the real world. In this essay, I offer a defense of the ideal/nonideal theory distinction according to which it amounts to nothing more than a division of labor, and explore some scientific analogies. Rawls's own focus on the ideal part of the theory, I argue, stems from a felt need to clarify the foundations of justice, rather than a utopian neglect of real world problems.*

KEY WORDS: idealization, theoretical models, science, Rawls, Sen, David Schmidtz, Justice, ideal theory, nonideal theory, physics, Ideal Machines, Ideal Pendulum

Rawls ignited a debate in political theory about how much compliance can be legitimately assumed in constructing a theory of justice. The roots of the debate lie in Rawls's distinction between ideal and nonideal theory. He writes:

> The intuitive idea is to split the theory of justice into two parts. 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. My main concern is with this part of the theory.[1]

* I would like to thank Jacqueline Ismael for tremendously helpful comments on an early draft. Paul Bloomfield provided extensive commentary that made the essay much better, and much less naïve, than it would have otherwise been. I would also like to thank the students in David Schmidtz's political theory seminar, and Michael Gill, Rachana Kamtekar, Geoffrey Sayre-McCord, Gerald Gaus, and other contributors to this volume for very stimulating discussion, especially Alexander Rosenberg who provided deeply insightful commentary on my essay. Most of all, I owe a deep and hearty thanks to David Schmidtz who invited me to contribute to the discussion, against my own better judgment, but with the effect of including me in a conversation that has been very rewarding for me.

[1] John Rawls, *A Theory of Justice*, rev. ed. (Cambridge, MA: Harvard University Press, 1999), 216. Partial compliance theory is to be devoted to ascertaining "how the ideal conception of justice applies, if indeed it applies at all, to cases where rather than having to make adjustments to natural limitations, we are confronted with injustice."

doi:10.1017/S026505251600039X

The idea is to form a positive conception of justice, in a setting that assumes perfect compliance, and take up the question of what happens under departures from compliance in a different part of the theory. As he says

> When we ask whether and under what circumstances unjust arrangements are to be tolerated, we are faced with a different sort of question. We must ascertain how the ideal conception of justice applies, if indeed it applies at all, to cases where rather than having to make adjustments to natural limitations, we are confronted with injustice. The discussion of these problems belongs to the partial compliance part of nonideal theory.[2]

The assumption is made not because it is held to be true, or approximately true, but to fix ideas about what justice looks like when everyone is acting as he or she should, in one part of the theory and to address issues about how much compliance can be expected, and how to respond to noncompliance, separately, in another part of the theory.

In what follows, I am going to be exploring some scientific analogies. I will begin with the rationale for focusing on ideal theory. Then I'll say few words about idealization in science and introduce the analogies that strike me as illuminating. I will go on to look at some of the objections that have been leveled against ideal theory and use the analogies to offer responses. I will argue that Rawls's focus on the ideal part of the theory stems from a felt need to clarify the foundations of justice, rather than a utopian neglect of real-world problems. Although I will rely on Rawls's own remarks about ideal theory, I will not limit myself to them. The goal will be to offer a *Rawlsian* case for ideal theory.

## I. The Need for a Theory of Justice

Rawls thought that there was a special need for a systematic understanding of justice. Why do we need an ideal theory of justice according to Rawls? To help clarify the concept, link it to pre-theoretic ideas about fairness, and offer explicit principles to assign basic rights and duties and to determine the division of social benefits. Our ideas about justice are emotionally charged, but confused and inchoate. Ideal theory articulates the justification by tracing it back to a conception of fairness in the original position and provides explicit principles for the design of institutions for regulating their claims against one another. Rawls undertakes to show in *A Theory of Justice* how the explicit principles follow from original agreement in a situation of equality. He takes his theory to be justified both by the conception of fairness embodied in the original position and the agreement of its implications with pre-theoretic intuitions about the

[2] Rawls, *Theory of Justice*, 309.

just distribution of benefits. Showing how the less intuitive derived principles[3] follow from rational choices in the original position is supposed to lend support to those principles, and to the consequences of those principles in disputed cases.

The ideal part of the theory in Rawls's system was devoted to the construction of a model of a just society in a setting of perfect compliance. Questions about how to deal with noncompliance were separated for the purposes of fixing the core content of the concept and clarifying its links to fairness. That concept would then be deployed in settings where the presence of other factors makes its expression more complex.

The ideal part of the theory *organizes the demands of justice* around the idea of rational choices made in the original position in something like the way that a physical theory organizes its empirical consequences around a compact set of principles embodied in the laws. In doing so, the ideal part of the theory: (i) displays the content of "justice" in a purified setting, (ii) derives its implications for the basic rights and duties of the individual and the design of institutions, and (iii) articulates its justification by showing it to be derived from rational choices that would be made in an original position of equality.

The point of the project is to systematize our understanding of what justice is, and to make it precise enough to form the basis of a set of organizing principles for society. The ideal part of the theory grounds the less intuitive principles in a recognizable kind of fairness, and (thereby) makes its justification transparent. One may object to the conception of equality embodied in the original position, but if the rest of the development of the theory is correct, you know where to aim your disagreement.

## II. Idealization in Science

Because of the widespread use of idealization across the scientific disciplines, philosophers of science became interested in understanding its role in scientific practice. If one were looking to the philosophy of science for discussion of idealization, one would find a large and somewhat disorganized literature containing a great deal of discussion of models or explanations of phenomena that make assumptions about the systems to which they are applied that are known to be false.[4] The question in those

[3] There are intermediate derived principles, for example, that each person is to have an equal right to the most extensive basic liberty compatible with similar liberty for others, and more specific corollaries.

[4] Landmarks in this literature include: Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Oxford University Press, 1983); Ronald Giere, *Explaining Science: A Cognitive Approach* (Chicago: University of Chicago Press, 1988); Ernan McMullin, "Galilean Idealization," *Studies in History and Philosophy of Science* 16, no. 3 (1985): 247–73. Cartwright's book set off a firestorm of discussion. It is a collection of essays that argues that practices of idealization provide arguments for causal entity realism. Giere's book is a discussion of the role of

cases is how the models can be good models of real systems if they explicitly *mis*represent those systems, or how explanations can be good explanations if they explicitly make false assumptions. Philosophical controversy has been centered on questions about the nature and legitimacy of idealization. The view of most scientists about the use of idealized models to represent real systems is pragmatic and pluralistic. There are various kinds of legitimate and useful types of idealization. The justification for them is practical, relative to representational goals, and depends quite specifically on the details of the case. Idealizations in models used to represent real systems can be useful, but they can also go badly wrong. There are some cases of uncontested illegitimacy, and many cases of contested legitimacy. As interesting as this literature is, it is of limited value for the purposes of understanding the role of ideal theory in Rawls's political philosophy. If we want to look for scientific analogues of Rawlsian ideal theory, we should not be focusing on the representational uses models. We should be looking at nonrepresentational uses: places where scientists produce models that represent behavior under pure (or "ideal") conditions, without any illusions of *representing the actual world*.

## III. Newton's Ideal Pendulum

In Book I, Propositions 51–52 of the *Principia Mathematica*,[5] Newton introduces the term "corpus funependulum" to refer to what we would call a "simple pendulum," or "ideal pendulum." This is a weight suspended on the end of a massless cord suspended from a pivot. An ideal pendulum experiences no air resistance. There is no friction in the pivot, and there are no exogenous influences. There are two dominant forces acting upon a pendulum weight at all times during the course of its motion; the gravitational force pulling it toward the center of the Earth, and the tension in the string pulling it upward toward the pivot. Newton derives an equation for the behavior of an ideal pendulum (known as the Law of the Pendulum) that relates the period of the pendulum to its length and the strength of the gravitational field:

models in science that includes an argument for idealized models as realistic representations of empirical systems. McMullin's paper uses historical examples to explore the epistemic implications of a type of idealization traced to Galileo. There were some early attempts to draw general lessons about science from the use of idealizations, but pluralism and pragmatism about the many different uses is more characteristic of recent work. See, for example, Peter Godfrey-Smith, "The Strategy of Model Based Science," *Biology and Philosophy* 21 (2006): 725–40; Richard Levins, "The Strategy of Model Building in Population Biology," in E. Sober, ed., *Conceptual Issues in Evolutionary Biology* (Cambridge, MA: MIT Press, 1966), 18–27; Michael Weisberg, "Three Kinds of Idealization," *Journal of Philosophy* 104, no. 12 (2007): 639–59; and Newton da Costa and Steven French, *Science and Partial Truth* (Oxford: Oxford University Press, 2003).

[5] Isaac Newton, *Philosophiae Naturalis Principia Mathematica* (London, 1687).

$T = 2\pi\sqrt{L/g}$, where T is the period, L is the length of the pendulum (in meters) and g is the strength of the gravitational field.

Newton was under no illusions that actual physical pendula fit this description. He coined the term partly to distinguish his ideal pendulum from the material ones discussed in Book 2,[6] where he describes experiments with real pendula that he uses to explore the resistance of fluids. To understand why he bothers talking about ideal pendula, we need to look to the larger purpose of the *Principia*. The book was intended to unify celestial and terrestrial mechanics by showing that the motion of planets was due to the same force as the fall of an apple from a tree. It was intended, that is to say, to provide a unified theory of gravity. The book is organized axiomatically. After the preface and definitions, Newton presents his three laws. The rest of the book presents the mathematical and philosophical development of the theory, drawing out consequences of the laws and definitions. At the point in the *Principia* where he introduces his corpus funependulum, he was showing that the independence of the period of a pendulum from its mass follows from the first two of his laws. This result agreed with what was known about the behavior of actual pendula since Galileo's time, so its derivation from the laws lent important initial support to the laws, and it would play an important role in the derivation of further results.

But the primary reason that Newton talked about ideal pendula is that he was interested in *gravity*. For his purposes, air resistance, friction, and the inevitable presence of exogenous influences were all distractions. They introduced complications that made the equations of motion for actual pendula much more complicated and obscured the content of the gravitational laws. The effect of gravity on the motion of a mass close to the surface of the earth could be isolated and precisely characterized by suppressing these exogenous factors, but if they were included, they introduced terms whose values varied from case to case and obscured the common element due to the effect of gravity. Ideal pendula play both a theoretical role and an expository role. The theoretical role was that the derivation of the ideal pendulum law is an important stepping-stone in the development of the theory, which is later used to derive planetary motion. The expository role is that they help fix ideas about what gravity is. For if you want to know what gravity *is*, you need to know what gravity *does*, and the ideal pendulum provides an especially pure and easily visualizable illustration of the effect of gravity on the movements of a massive body. Newton's practice is followed to this day in physics textbooks and classrooms. The ideal pendulum law remains one of the first results derived from Newton's laws of motion, and the ideal pendulum remains a useful way of visualizing the effect of gravity.

[6] Newton, *Principia, General Scholium*, at the end of sec. 7.

When Rawls wrote *A Theory of Justice*, and focused on the ideal part of the theory, he did not set out to produce what purported to be a model of the actual world and *fail* any more than when Newton described the behavior of an ideal pendulum, he set out to describe the behavior of an actual pendulum and *failed*. In both cases, they were using the idealized model to express a part of the content of their theory. In Newton's case, it was to exhibit the effect of gravity on the motion of an object close to the surface of the earth in a pure form, unobscured by the presence of other forces. It was also to show how that effect follows from his first principles (namely, his three laws of motion). In Rawls's case it was to exhibit what a just society looked like in its purest expression, unobscured by noncompliance. It was also to show how the principles that organize such a society follow from the account of justice as rational choices made in an original position of equality. The full development of a complete theory of justice should have the resources to deal with the effects of noncompliance, but the ideal part of the theory is the part of the theory that most clearly displays the *content* of, and *justification* for, his notion of justice.[7]

There are other examples of the use of models of ideal systems in science that play a similar expository and theoretical role. The study of mechanics begins by learning about the behavior of ideal machines. These are nonactual mechanical systems (for example, pulleys, levers, crank and piston assemblies, wheel and axle systems) in which energy is not dissipated through friction, deformation, wear, or other inefficiencies. In relativity, we learn about the behavior of ideal clocks and measuring rods. These are systems that perfectly measure proper time and spatial intervals, never wearing out, running out of energy, or suffering the bumps and scratches that plague our own watches and rulers. In thermodynamics, one learns about the behavior of ideal gases. These are gases whose molecules occupy negligible space and undergo no interactions. Models of ideal systems, employed in this mode, are not *mis*representations of actual systems. They are not typically being used to *represent* real systems at all. They are used to teach something about what the theory says. In mechanics, we focus on ideal machines to assimilate the principles of statics and mechanics. In relativity, we focus on ideal rods and clocks to operationalize notions of space and time. In thermodynamics, we talk about ideal gases to focus on global dynamics. Ideal models allow us to isolate certain relationships, suppress complicating factors that we are not interested in, and explore in isolation features of the world that always come together in practice. What is being modeled in these cases are actual forces or laws in non-actual settings, that we find to be revealing for various purposes.

[7] The terminology of "ideal and nonideal *theory*" is a little misleading. It is more accurate to talk of ideal and nonideal models, or the ideal and nonideal *parts*, of the theory of justice. There is only one theory, but it has parts that focus, respectively, on idealized and nonidealized systems.

Note that when talking about ideal machines, ideal gases, or ideal pendula, the word "ideal" would be misleading if one took its evaluative connotations seriously. An ideal pendulum is not a "perfect" pendulum in the sense that it is a particularly *wonderful* pendulum, a *utopian* pendulum, and the kind of pendulum that we should all *hope* and *strive* for. The sense of "ideal" in play here is the one that contrasts with "real." An ideal pendulum is one that simply suppresses factors that are present in real pendula.[8]

There is no presumption that the factors that are suppressed in ideal models are *unimportant* for the purposes of understanding real systems. In some cases, the factors suppressed do not make much of a difference to the behaviors that we are interested in, so the conclusions that we draw about the ideal carry over to real systems. But in some cases, they make a big difference. Indeed, in some cases, we suppress particular factors because when they are present, they dominate, so the only way to understand *non*dominant influences is to suppress them.

Finally, note that the focus on ideal systems does not mean that we are not *ultimately* interested in real systems. It is rather that we work our way toward understanding what our theories say about real systems by understanding what they say about these simpler systems first. The theory of ideal machines makes this point clearly. Ideal machines suppress the effects of friction and wear and other ways in which energy is dissipated to the environment. Far from being unimportant, dealing with these inefficiencies is the defining problem of mechanical engineering. Even if we have a purely practical interest in making engines, however, we learn about ideal machines because one does not have a good understanding of the mechanics of real engines unless she has a good understanding of the mechanics of ideal ones. Understanding ideal machines is part of understanding the more complex reality of actual ones.

People sometimes make the mistake of thinking that the fact that no real pendulum behaves exactly like an ideal pendulum shows that Newton's laws are only *approximately true* of actual pendula. That is incorrect. Assuming that we live in a classical world, every system is modeled with perfect accuracy and precision by the Newtonian laws.[9] But the form that the laws would take for actual pendula is more complicated, because actual pendula are subject to other forces and exogenous influences that vary from one to the next. The correct thing to say is that actual pendula only approximate the *simple form of the laws* exemplified by an ideal pendulum. Newtonian models of real pendula are much more complex. That complexity makes them better at representing real systems, but

[8] An ideal pendulum is a perfect exemplar of simple harmonic motion, so it is in that sense "ideal," and an ideal machine is maximally efficient, so it is in that sense "ideal," but "idealization" in its most basic meaning here is simply the suppression of factors that are present in real systems.

[9] Of course, we do not live in a classical world, but that will not matter for our purposes. The reasons that we do not live in a classical world do not affect the points made here.

worse at conveying the contribution of gravity, because the contribution of gravity is obscured by the presence of other factors.

## IV. Theorizing: How and Why

In the presentation of a physical theory, the laws and the quantities that are part of the theory are introduced together and expressed as first principles. The development of the theory shows how all of the complicated movements of actual objects can be seen to flow from them. This allows the organization of all of the motley motions of material bodies around three simple laws. The theory is justified (to the extent that it is justified) by the fit between the consequences and the phenomena. The actual process of coming up with a theory is a constant movement back and forth, proposing first principles and adjusting them to get the right fit between their consequences and the phenomena. We *craft* our first principles so that they allow us to derive consequences that match the phenomena, and then we *use* our first principles to derive predictions about what will, or would happen, under conditions that have not been observed. So, the theory gets *justified* by showing us how to recover phenomena that have been observed, and then *applied* to derive new phenomena.

There's a natural analogy with the reflective equilibrium that Rawls describes as the process by which one arrives at a theory of justice. The first principles in that case are not laws, but a conception of justice. In Rawls's theory, that conception of justice is captured in the idea of rational choices from an original position of equality. The theory gets *justified* by showing us how to recover judgments about intuitively clear examples of injustice, and then *applied* to adjudicate grey cases and deliver principles for constructing institutions.[10]

[10] There are differences between descriptive and normative theories, but they do not affect the epistemic analogy drawn here. In both cases we have a set of first principles, consequences are drawn from them and compared against a stock of beliefs obtained from an independent source. And in each case, the theory is judged by how well it does reproducing the stock of independently sourced beliefs.

One reason for suspecting that there is an important disanalogy may stem from an overly simplistic view about how theoretical terms in science get their reference. One might think that in the case of gravity, there is a thing out in the world that our various theories of gravity are trying to characterize correctly. Our theories go wrong by mischaracterizing the behavior of that thing. The standards that govern correctness, in that setting, are independent of the theory and independent of any choice or definition on our part. Whereas in the case of justice, we are presented with theories that introduce different conceptions of justice as a definition, and there is no fact of the matter, independent of the standards for accepting a theory, about whether the theory gets it right.

This disanalogy is illusory. Gravity is a theoretical concept introduced by a theory that systematizes motion. The everyday idea that gravity is what pulls things toward the center of the earth gives the concept a little pre-theoretic content, but not much. When you accept a set of laws into which gravity enters, you accept a definition of what gravity is, in something very like the way that when you accept a theory of justice you accept a definition of what justice is. In neither case is there a fact of the matter, independent of the standards for accepting a theory (at least none that plays a role in science), about whether the theory gets it right.

There is a practical reason for wanting a theory of gravity: to form a clear and distinct idea of the effect of gravity on motion that allows the expression of precise laws. The laws can be used not only to predict, but also to intervene effectively in nature. There is a more theoretical reason for wanting a theory of justice: to clarify the foundations of the concept, because our pre-theoretic ideas about fairness are too unsystematic and equivocal to serve as a basis for the design of institutions. Think of the child here whose outrage at her sister getting more is quieted by parents who explain that she got more last time, or that she will get as much as her sister when she is her sister's age. Our ideas about fairness are, in that sense, both equivocal, and educable. They are equivocal because there are many different ways of gauging equality in any given situation. And they are educable because we can be persuaded that our pre-reflective judgments of inequality employ the wrong standard. A theory of justice shapes those pre-theoretic intuitions into something systematic and precise to provide (as Rawls says) the foundation charter for a society.

## V. What is the Rationale for Idealizing Compliance?

Idealizing assumptions are always specific in their content. What we ignore and what we attend to depends on what we are interested in showing, expressing, or exploring. So, for example, in the case of ideal pendula, Newton ignores air resistance and friction, because he is interested in exhibiting the effect of gravity. In the case of ideal machines, we ignore sources of inefficiency because we are interested in conveying the principles of statics and kinematics. In the case of ideal gases, we ignore the interactions among molecules and their spatial volume, because we are interested in the global dynamics. The ideal theory in Rawls is not ideal in every respect. He makes all kinds of realistic assumptions about what he calls "the fixed constraints of human life," including, for example, that there is neither an overabundance of goods nor severe scarcity. The rationale for these two assumptions is one that he took from Hume, namely, that in case of overabundance, there would be no need for principles of fair distribution, and in cases of severe scarcity, the principles of justice would be (as Hume puts it) "suspended" in the interests of self-preservation.[11] The specific respect in which he idealizes is by suppressing something that is always present in our world: noncompliance. The assumption of perfect compliance is acknowledged to be unrealistic. Rawls is perfectly aware that human behavior is motivated by many factors, and that perfect compliance with the principles of justice is not a realistic expectation. He recognizes that things like anger, love, strategic self-interest, and myriad

[11] Hume says that they will be suspended in fact. It is not clear whether he thought they should be suspended in principle, or whether the question is one that he would have recognized as sensible.

other forms of partiality play a role in practical reasoning, and they often weigh against the demands of justice. What is the rationale for focusing one's energy on understanding what justice would look like in a setting in which there was perfect compliance, if perfect compliance cannot be realistically expected?

Rawls bracketed noncompliance, in the way that Newton bracketed air resistance and friction: namely, not because they are negligible or to be disregarded, but because he saw a need for clarifying the foundations of the concept. The idealized model conveys the content of his conception of justice, and exhibits the connection between justice and fairness in the clearest and most transparent way.[12] Since the connection between justice and fairness motivates the principles of justice across all contexts, including their expression in settings where there is noncompliance, it plays an important heuristic role. An analogy with dividing a dinner bill is helpful here. If you are presenting rules for dividing dinner bills, it makes sense to start with the case in which everyone is doing his or her part, because it is in that setting that the ideas of equality and fair share that guide the division have their simplest and most transparent expression. The more complex rules that apply to settings in which some people have left without paying can then be motivated by reference to the pure case, by showing how fairness becomes complicated by noncompliance.

There is nothing in this division of labor that suggests that facts about how likely people are to comply with it should be *ignored*, or that the effects of noncompliance are negligible. It is simply an attempt to clarify the foundations of the concept in a setting in which the connections to rational choice in a position of equality are most transparent. Taking noncompliance into account clouds those issues, though it is of acknowledged practical importance for the purposes of actually building a (more) just society. The conceptual effort of clarifying the content of his conception of justice and exhibiting its justification is one that is motivated, in Rawls's mind, by the fact that our ideas about justice are too confused, equivocal, incomplete, and too concretely tied to emotions, to provide principles for constructing institutions. But they are also fundamental to the foundations of decent society.

## VI. Objections to Ideal Theory

The assumption of perfect compliance has encountered resistance for various reasons. First, because how much compliance is to be expected depends on what the principles of justice are. It is, that is to say, an

[12] One might fairly wonder why models of justice in a setting of full compliance should have a privileged role fixing the content of justice, if full compliance cannot be expected. That is not quite the right way to think of it. The right way to think of it is not so much that the idealized model has a privileged role *fixing* the content of justice, as that it *displays* that content in a particularly clear way.

endogenous variable, and it has been argued that a conception of justice that places unrealistically high demands (demands that nobody would comply with) cannot be workable as a society for human beings. That is certainly correct. Let us suppose that it is a well-defined question how well some particular conception of justice will fare once the facts of noncompliance are taken into account. In asking how much compliance can be legitimately assumed in constructing a theory of justice, we consider the degree of compliance that can be expected after suitable socialization.[13] That is something that can and should be taken into account in evaluating institutional designs as things that we should try to implement. But to build the expectation of noncompliance into our concept of justice is like building into our conception of the correct cooking time for a soufflé, a correction for the fact that we always overcook. We might reasonably build the correction into our cooking *instructions*, which are formulated to optimize the resulting behavior; but to build it into our idea of the correct cooking time is to mischaracterize the concept. If we have a conception of what the correct cooking time is, guidance for the formulation of instructions should be obtained by comparing the *correct* cooking time with information about how we tend to miss it. Having a concept of the correct cooking time is not always necessary. We might be able to get by with instructions crafted to bring about the right result. But the extra articulation is desirable because it helps us become better cooks. It allows us to improve, that is to say, and also provides a flexible schema for generating instructions that bring about the right result for people with different tendencies.

One might still wonder how we should *evaluate* theories where the demands of compliance are too high to have any real expectation of being fulfilled? There are really two questions here: (1) How should we evaluate theories of justice *as conceptions of justice?* and (2) How should we evaluate them as solutions to the practical problem of designing institutions? I will discuss these questions in turn.

(1) Can it be justice if it demands more than human beings will realistically do? We can make perfect sense of the idea that people are not generally as just as they should be. Or even that they are not *typically* as just as they should be. Justice can (and should) demand more of us than we tend to give anyway. It can (and should) be aspirational. But can the requirements on individual behavior in a just society be so strong that they make compliance very unlikely, people being what they are? Let's ask the same question about other concepts. We have relatively well-defined conceptions of altruism, courage, and cruelty, and it is arguable that we

[13] This allows us to pass over subtleties about whether we should be thinking of the degree of expected compliance as merely a *fact*, or a necessity grounded in human psychology. We leave open what counts as "suitable" socialization, though presumably it should be neither coercive nor overly costly.

get an especially pure expression of what altruism, courage, or cruelty *is* by seeing how people would act if everyone acted only out of that motive. The fact that nobody might realistically behave that way does not stand against these hypothetical results as expressions of the content of altruism, courage, or cruelty.

Here, I am echoing David Estlund's elegant defense of the claim that a realistic expectation of compliance might be a requirement "on any conception that is a serious candidate for implementation, but it is hard to see how it could be part of the notion of *justice*."

And again,

> Surely, society should not implement institutions that people will not be able to bring themselves to comply with . . . . The question is whether that is a constraint on the content of justice. The rules and institutions that should be constructed given what is known about everyone's likely compliance are hardly guaranteed to be rules and institutions that qualify a society as just.[14]

But how far does this go? Could it be *justice* if it demanded more than any (or most) of us could — in a suitably strong sense of "could" — give? Does a theory of justice fail to capture the content of justice if it entails that real people always (and perhaps inevitably) fall short of its requirements? Perhaps it is somehow implicit in the notion of justice (as distinct from altruism or courage, for example) that perfect justice must be — by its nature — attainable for an ordinary human being. Perhaps, for example, (x is just) → (x can be reasonably demanded of me), and (x can be reasonably demanded of me) → (I *can* x), in some suitable sense of "can." This kind of requirement is implicit in the comment that people often make about morality when they say that we should not be looking for a theory of morality for angels, but a theory of morality for humans.

This is a delicate issue. One might argue that we get a better understanding of the concept of what justice is if we can see justice for humans, justice for Martians, and justice for angels, as all recognizable as forms of *justice*. In that case, justice for humans would emerge as a special case of a general concept, obtained by seeing how human limitations shape its content.[15] For our purposes, however, this is not a point we need to press. There is nothing unrealistically demanding about what compliance demands on Rawls's conception of justice, and he was quite concerned

[14] David Estlund, "Human Nature and the Limits (If Any) of Political Philosophy," *Philosophy and Public Affairs* 39, no. 3 (2011): 226. I am indebted to Estlund's probing and careful discussion.

[15] Science, or at least physics, characteristically seeks this kind of articulation. A natural scientific analogy here is the relationship between the Special and General Theories of Relativity. The Special Theory emerges as a special case of the General Theory, obtained by setting the curvature tensor to 0.

that it not do so. He would have seen it as a defect of his conception of justice if it were not realistically attainable for human beings.[16]

(2) The second question — How should we evaluate theories where the demands of compliance are too high to have any real expectation of being fulfilled as a practical solution to the problem of designing institutions? — is easy to answer. A theory of justice would be a terrible solution to the practical problem of designing institutions if the expected outcome of trying to implement it is far from the ideal. Solutions to practical problems should *optimize expected outcome*. The connection between the expression of the ideal and the expected outcome of an attempt to implement it is by no means direct. A theory that has no realistic hope of successful implementation may be not worth aiming at for various reasons, not only because there is no hope of getting there, but because aiming at an unattainable ideal is not guaranteed to be a good way of approaching that ideal. Indeed, aiming to realize such a theory may even produce worse results than we currently have.[17] We can make this point with the example of ideal machines. An ideal machine is one that exhibits maximal efficiency. If we construct a real-world engine on the model of an ideal machine, we will inevitably construct one that does not work. An ideal machine does not lose energy through dissipation into the environment. Any real engine does. This simply emphasizes that *non*ideal theory is essential and ineliminable in producing solutions to practical problems. Whether the ideal is something that we should implement or aim at depends in detail on facts that are not internal to the ideal theory.

This is a place where the word "ideal," and its history in political philosophy, might have a misleading and pernicious influence. It strongly suggests something to be aimed at. There is less tendency in the scientific examples to be misled in this way, but perhaps Rawls would have done better to choose a different word. It is not clear whether he himself was entirely clear on the matter. So, for example, in *Law of Peoples* he writes, "until the ideal theory is identified . . . nonideal theory lacks an objective, an aim, by reference to which its queries can be answered."[18] This has been taken by some to suggest that ideal theory is required as a target for steps in the right direction, and has consequently sent people along a path that turned out to be a dead end. In retrospect, it is easy to say that Rawls should have been rather firmer about the distinction between an ideal conception of justice as (i) a gauge for how just a society is, (ii) the model or template on which an actual society is to be built, and

[16] Thanks to Michael Gill for this observation.

[17] These were points that were very effectively made by Amartya Sen, *The Idea of Justice* (Cambridge, MA: Harvard University Press, 2009); and David Schmidtz, "Ideal Theory: What It Is and What It Needs To Be," *Ethics* 121 (2011): 772–96 in response to Simmons's attempt to defend Rawlsian ideal theory as necessary in order to rectify injustice: A. John Simmons, "Ideal and Nonideal Theory," *Philosophy and Public Affairs* 38 (2010): 5–36.

[18] Rawls, *Theory of Justice*, 90.

(iii) something to "aim at" in taking steps to make an unjust society more just. He should have endorsed (i), but not (ii) or (iii). He should have abandoned any suggestion that the ideal theory provides something to aim at. And he should have emphasized the content-defining and justification-displaying role of his theory of justice, that is, its role articulating the content and consequences of his conception of justice.[19]

Considerations of compliance should be "set aside" only for the purposes of defining the concept of justice. They need to be addressed explicitly and systematically in the nonideal, practical part of the theory. The inevitable inefficiencies of actual engines are set aside only for the purposes of assimilating the principles of statics and kinematics. They are addressed explicitly and systematically in the practical part of mechanics. It is certainly true that, as David Schmidtz and others have emphasized, solving idealized problems does not generally yield approximations of solutions to real problems. But we can motivate ideal theory without making that mistake by insisting on the ineliminability of the nonideal part of the theory for the purposes of identifying practical solutions, and giving the ideal theory a different role. The rationale for ideal theory is only the rationale for a certain division of labor, one that allows the clear, explicit articulation of a concept of what justice demands in one part of the theory, and separate consideration of what to do when those demands are not met in another.

The separation of the ideal and nonideal parts of a theory is one that comes from within a theory, and often emerges only as the theory matures. Consider again Newton's *Principia*. It was a considerable achievement for Newton to be able to formulate precise laws for the effect of gravity on the motion of a pendulum. Since the behavior of a real pendulum is always more complicated than the ideal pendulum, in order to isolate the contribution of gravity, he had to effectively solve for the effects of friction, air resistance, and exogenous influences *at the same time*. Only once those forces are themselves understood does the effect of gravity emerge clearly and distinctly in a form that allows precise characterization. The separation of the contribution of gravity to the movements of a real pendulum from the effects of these other forces is only a virtual separation, since in practice they always come together. But it is a huge theoretical achievement. When Newton presents his theory, he gives the ideal part of the theory first — that is, the part that describes the effect of gravity alone, and the simple, precise laws that describe that effect — because the achievement of the theory is to isolate its contribution.

To the extent that the purpose of a theory of justice is to articulate our moral ideas enough to separate justice as a distinctive notion — that is, to say what claim justice has on the design of our institutions, and our

[19] Laura Valentini, "Ideal versus Nonideal Theory: A Conceptual Map," *Philosophy Compass* 7 (2012): 654–64.

individual behavior (as against, for example, empathy or etiquette) — it makes sense to strive for such a theory.[20] And it also makes sense to give it priority in the presentation and teaching of the theory of justice. The very same considerations, however, caution against thinking that the ideal part of the theory is complete, or can stand alone. As with Newton's model of the corpus funependulum, or the theory of Ideal Machines, it can be related to the actual world only in conjunction with a nonideal part.

## VII. The Need for Practical Solutions to Practical Problems

This does not settle the issue entirely because we can recast the worry as a concern that looking at justice under conditions of perfect compliance does not give us a very interesting notion of justice, because it eliminates the problems that a conception of justice is needed to solve.[21] It is like coming up with a theory of how to fly in the absence of gravity. Where there is no gravity, we do not need to *fly* to remain airborne. It is only because of gravity that flying is needed. And just so, one might suppose, in saying what justice looks like in a setting in which everybody is complying with the demands of justice, Rawls does not solve any of the difficult problems. The difficult problems all lie in the part of the theory that Rawls sets aside, namely, in cases where (as he says) "we are confronted with injustice." An analogy with conditions of overabundance can be used to motivate the worry. Rawls was well aware of Hume's famous remarks on that topic. "Let us suppose," Hume says,

> that nature has bestowed on the human race such profuse abundance of all external con-veniences, that, without any uncertainty in the event, without any care or industry on our part, every individual finds himself fully provided with whatever his most voracious appe-tites can want, or luxurious imagination wish or desire . . . . No laborious occupation required: no tillage: no navigation. Music, poetry, and contemplation form his sole business: conversation, mirth, and friendship his sole amusement. It seems evident that, in such a happy state, every other social virtue would flourish, and receive tenfold increase; *but the cautious, jealous virtue of justice would never once have been dreamed of*. . . .

The idea here is that the difficult and interesting problems that principles of distributive justice are needed to resolve arise only in conditions of scarcity. Where there is abundant surplus, everybody can have whatever he or she wants, and justice is trivial. Just so, one might say, the difficult and interesting problems for political theory are problems involved in

[20] Although Rawls's theory is, in the first instance, a theory about the design of institutions, there is a connection to individual justice: knowing what would count as complying with the demands of justice tells us what justice demands of each of us.

[21] See Jacob Levy, "There Is No Such Thing as Ideal Theory," this volume.

confronting injustice. This is a concern that Schmidtz has pressed with special persuasiveness. He writes:

> We can and must set aside distracting details and focus on the problem — on the human condition insofar as we are theorizing about politics or justice — even though any characterization invites accusations of begging someone's version of the question. But one thing we must not set aside as a detail is the problem.[22]

And again, later, "We can go badly astray if we strive for what Rawls called a 'systematic grasp of more pressing problems' by assuming away those very problems."[23]

Let us consider, again, the theory of ideal machines on which every beginning mechanical engineer cuts her teeth. These are machines that idealize away the defining difficulty of making engines, because they do not dissipate energy through wear, or friction, or heat. Making engines is all about how to deal with the fact that real machines are inefficient as a matter of physical law. The example shows clearly that models of engines that exhibit maximum efficiency do not tell us how to build an engine that actually works, the world being what it is. Nor do ideal models tell us how to build engines that are *more* efficient than the ones we have. Difficulties arise in the difference between ideal and nonideal machines that take all of the resources of the nonideal part of the theory to resolve.

One can understand the objection that Schmidtz is raising as the charge that all of the *important* problems for political theory arise in the difference between the ideal and the actuality. This leads directly to one of the most influential complaints about ideal theory: that it does not address practical problems, that is, that it is an intellectual exercise with little *practical* value. This complaint does not challenge the thought that ideal theory plays a role in defining the concept of justice. It simply questions the importance of that role. This evaluative judgment is explicit in some places in Schmidtz's writings. He says, for example: "If anything needs to be set aside and treated as a mere distraction from *work worth doing*, it is visions of how well a system would work but for the recalcitrant reality of human beings."[24]

I have said why I think there is a more sympathetic way of understanding what Rawls is doing. He is not downplaying the need for practical work, but advocating a division of labor that makes room for the theoretical work of clarifying the foundations of the concept of justice. One might

[22] David Schmidtz, "Ideal Theory," unpublished manuscript. I'm very grateful to David for showing me his manuscript in draft and allowing me to quote from it. The manuscript is currently in press and due to be published in *Oxford Handbook of Distributive Justice*, ed. Serena Olsaretti (New York: Oxford University Press, 2016). All quotes are from the unpublished manuscript.

[23] Ibid., 1–2.

[24] Ibid., 21.

think that this is work worth doing in its own right. But for someone who thinks that purely theoretical work has no intrinsic value, it is worthwhile saying why it is not so easy to separate the practical and theoretical. Rawls himself clearly thought that there was a *practical* need for the theoretical work, and the scientific examples support that suggestion. Consider again the theory of ideal machines. Ideal machines would serve very poorly, if deployed as models for building an efficient engine, because real engines wear out and break down; they dissipate energy into the environment. The practical difficulty of making engines is *all about* trying to overcome these inefficiencies. So the theory of ideal machines is worthless on its own for the purposes of making engines that will work. The ideal theory, however, does not have to function *on its own*. It is an ineliminable *part* of a theory that has a nonideal part as well. The two come together as part of a package deal, and they work together to generate practical solutions to real-world inefficiencies. An engineering student who left school with the theory of ideal machines under her belt will have *only* the first step in an education that leads to practical knowledge, but she *will* have that first step. And her education would not be complete without it. The nonideal theory builds on the ideal theory, adding the factors that pull against the ideal, introducing the problem of inefficiency and also supporting instrumental reasoning about how to deal with it. The theory of ideal machines is part of an articulated theory that defines maximal efficiency, identifies sources of inefficiency, and provides the theoretical knowledge needed to address these inefficiencies.

Is it possible to do engineering without a theory of ideal machines? Yes. People learn to make things without this sort of articulated understanding all the time. But they do not always do it as well. This sort of articulated understanding makes us better at predicting, intervening, and designing systems. All of its other virtues aside, science is a very useful handmaid to engineering.

There is another way in which it is hard to separate the purely theoretical enterprise of clarifying the conceptual foundations of justice from the practical problems of making our societies (more) just. Understanding whether and in what sense the rules that regulate our societies are just can play a role in people's views about whether they *have a reason* to comply. To take the theoretical project seriously is to treat the citizens to whom the rules of justice apply with enough respect to think that clarifying the foundations of the concept of justice is worth doing. And that is — in its turn — just to recognize that how people behave is ultimately up to them. If considerations of justice are to mitigate people's strategic interests, having a clear presentation of its content and justification are essential. Getting people to comply should not be a matter of imposing rules, but of showing people that they have a reason to comply: providing a clear and convincing derivation of the rules of justice from a kind of fairness they can endorse.

It should also distinguish considerations of justice from kindness, altruism, civility, and other virtues. Knowing what distinguishes justice from these and other virtues gives it a different kind of traction in our practical reasoning. It would be nice if we were all kinder, more altruistic, and more civil than we are, but justice has a different kind of claim on our behavior that is revealed by the connection to fairness. Perhaps it is naïve to think that people care about justice, or to think that they care enough for it to impact their behavior. I don't believe that, but I also don't think it matters. The effort of clarifying foundations shows respect for the people to whom those principles apply.[25]

Schmidtz remarks that "Much of what we currently call ideal theory is an exercise in imagining how we would reinvent the world if only we could start with a clean slate and do a complete reset, rebuilding society from the ground up." That description need not apply to every exercise of ideal theorizing, although the word "ideal" used in this context, gets in the way because it suggests that "ideal theory" is a theory about what we would ideally do, or about "the ideal," rather than just the part of the theory that displays what justice looks like in a setting unadulterated by noncompliance. I have said why the fact that the ideal part of a theory does not provide practical solutions *on its own* does not mean that it is not a part of the task (and indeed, an *essential* part of *one way* of approaching the task) of providing solutions. It might be that political theorists have found this part of the theory more attractive and have neglected its practical component. And one might easily object to the choice to *start* with ideal theory in the order in which we actually set out trying to address the problems of the world.[26] One might well agree with Sen that there is enough manifest injustice in the world that our attempts to address practical problems should not wait for a solution to the theoretical problem of working out an ideal theory. To wait for a solution to the theoretical problem of working out an ideal theory before addressing practical problems would be like making the building of bridges wait until we have a final theory of physics. That is not how it happens in science, and there is no reason to think that it has to happen that way in political theory. The theoretical and practical parts can go on simultaneously and inseparably, each

[25] Jason Brennan and Geoffrey Sayre-McCord, "Do Normative Facts Matter . . . to What Is Feasible?" this volume) argue persuasively that normative truth matters to people and makes a difference to their practical reasoning.

[26] Rawls does seem to have thought that we should start with ideal theory, and when that is finished, proceed to the nonideal part. He writes in *Theory of Justice*, "Nonideal theory, the second part, is worked out after an ideal conception of justice has been chosen." This is a large part of what Schmidtz finds objectionable: "articulating ideals is not the right place to start; if we start with a problem, then our starting point has the potential to discipline our reflection on what to count as a solution" ("Ideal Theory," 21). The scientific examples support Schmidtz here. We start with the practical problem of making the world better; the ideal and nonideal parts of the theory develop together, as part of a package deal, judged by their joint capability to deal with real-world problems.

driving the other forward. Clarification of the conceptual foundations of justice — getting clearer and sharper on what justice is (and if not forging agreement, at least understanding our differences) — should be part and parcel of recognizing and addressing injustice.

Let me briefly review here some of the lessons of the discussion so far.

- In the scientific examples, there is no such thing as "ideal theory." There are models, rather, of ideal systems. These typically suppress factors that are present in real systems, or incorporate simplifying assumptions of other kinds. They are drawn from a theory that also has the resources to model nonideal systems.
- There is no general presumption that ideal models provide approximations to the actual case (something that is ill-specified, in any case, until we say what particular features are being approximated and the degree of precision in question). Sometimes they do, but sometimes they don't, and they are not offered as approximations to the actual case in the scientific examples that provide the most illuminating analogues to Rawls's use.
- It is misleading to talk about "idealizing assumptions" to the extent that this carries the suggestion that they are false assumptions about the actual world. We should speak rather of theoretical models of hypothetical systems. Theories aim for an articulated conception of their target domain, adequate to provide accurate models of real systems. Models of ideal systems allow us to explore in isolation, factors that always come together in practice, or to display a particular effect in a simplified setting.
- Idealizations are always specific in their content, and they typically serve a particular role. This can be useful in various ways, but they can also go wrong. Whether an idealization is appropriate depends specifically on the context and purpose. There is little of generality to say about what makes an idealization a good one. The only rule is that one should use discretion and care.

The division between the ideal and nonideal parts of a theory of justice is nothing more than a division of labor that separates the effects of noncompliance for separate treatment. It does not carry with it any an argument for neglecting the sort of practical engagement with real-world problems that authors like Schmidtz and Sen advocate.

## VIII. An Alternative to Ideal Theory

Schmidtz suggests an alternative conception of the project of political theory, one that does not demand ideal theory, but rather represents the task of political theory as crafting solutions to problems as they emerge. Schmidtz says: "Where there are facts, where facts are subject to change

in ways that matter, and where there is something we can do, we have a problem . . . Problems give us criteria for sorting out what to count as a solution."[27] I'm not sure that this can proceed entirely without ideal theory, as conceived above. The facts are always subject to change in ways that matter, and there is almost always something that we can do. An ideal theory of justice fixes the content of the concept in a way that allows us to identify injustice *as injustice* (rather than as simply a regrettable condition). That gives it a special status and a special claim on us to address as a society.

Sen has argued that there is more important work to do that does not demand a clearly articulated general conception of justice. To him, the point of theorizing about justice is to help us characterize and then undo *manifest* injustice. For those purposes, he argues, an ideal theory of justice is neither required nor useful. My own view is that this underestimates how much grey area there is. There are a lot of things that are manifestly *wrong* with the world. But how much of it, and what parts, are *unjust*?[28] Justice is a special concept with a special claim on public action to remediate. It is just as important to limit its demands as it is to articulate them. The centrality of the concept of justice, and the importance of its political function, give the need to clarify its foundations a special import. People's pre-theoretic conceptions of injustice are too thin a reed on which to hang political theory. Like the angry child mentioned earlier who has explained to her carefully how a division of favors (or a set of rules) that does not seem fair at first, may nevertheless *be* so, one might look to political theory to articulate and educate our pre-theoretic ideas about what is just. A case for ideal theory that survives these critiques is the one that looks to go beyond the cases of manifest injustice and articulate a concept that rules on the great grey area and also provides a justification for pre-theoretic intuitions about injustice.

## IX. A Worrisome Challenge

Schmidtz offers a more radical challenge to the idea that we should be trying to articulate a positive conception of justice at all. He writes "what if justice were simply an absence of injustice? In that case, seeking an essence of justice would be like seeking an essence of "non-dog."[29] This is an interesting suggestion, and it finds some support with a view of both scientific and ethical theorizing that Philip Kitcher has advocated. Kitcher, tracing the roots of his view to Dewey, argues that science and ethics should both be thought of as ongoing human projects without a clearly defined goal or endpoint. Schmidtz thinks that political theorizing should

[27] Schmidtz, "Ideal Theory: What It Is and What It Needs To Be," 4.

[28] See Judith Shklar, *Faces of Injustice* (New Haven, CT: Yale University Press, 1992) for the difference between injustice and misfortune.

[29] Schmidtz, "Ideal Theory: What It Is and What It Needs To Be," 3.

likewise be a matter of responding to new forms of injustice as they arise, and he is suggesting here that the attempt to form a positive conception of justice might be misguided right out of the gate. An analogy might be made here by thinking about medicine. The goal of medicine is a negative one: the elimination of pathology. The method for achieving it has to be piecemeal and adaptive. We have to identify and address diseases one at a time, as they arise, developing tools to fit them in the context at hand. We cannot know in advance what the perfectly healthy person would be like because the body adapts to changing circumstances, new diseases coevolve with it, and an adaptation that serves us well in one setting might undermine us in another.[30] For these reasons there is no well-defined concept of perfect health that could be fixed in advance and made the target of inquiry. Whether justice is like that I want to leave as an open question. This criticism is not just a criticism of ideal theory, but of theorizing about justice at all.

## X. Conclusion

From an outsider's perspective, the ideal/nonideal theory debate looks more like a dispute about what kind of theorizing is worthwhile doing than a competition between genuinely competing projects. The scientific analogies suggest that ideal and nonideal theory are actually deeply bound up with one another and that they can (and should) go on simultaneously. It would be wrong to think that either type of theorizing should be foregone, or that either has the kind of priority that would make progress on one wait on resolution of the other. As for dispute about where the important work for political theory lies, we can tolerate disagreement on that. The suggestion that it might be *a waste of time* to try to clarify the conceptual foundations of a concept that plays such an important role in public debate, in the construction of public institutions, and in the regulation of interactions among citizens, seems much too strong. I think that the most convincing take home lessons — very important ones, that perhaps needed to be made —is that political theory should no more be *only* about ideal theory, than mechanics should be *only* about ideal machines.

Defending a role for ideal theory, of course, leaves the substantive questions about the content of justice, and the design of institutions entirely open. One might not agree with Rawls's theory, but to object to the attempt to clarify the foundations of the concept, and express it in the form of an ideal theory that transparently exhibits its content and justification, strikes me as misplaced.

*Philosophy, University of Arizona*

[30] See also Alexander Rosenberg, "On the Very Idea of Ideal Theory in Political Philosophy," this volume.