

What's Ideal About Ideal Theory?

One of the main tasks that occupies political theorists, and arouses intense debate among them, is the construction of theories—so-called ideal theories—that share a common characteristic: much of what they say offers no immediate or workable solutions to any of the problems our societies face. This feature is not one that theorists strive to achieve but nor can it be described as an accidental one: these theories are constructed in the full knowledge that, whatever else they may offer, much of what they say will not be immediately applicable to the urgent problems of policy and institutional design. Since this may seem puzzling, and has been subjected to severe criticism, the main task of this paper is to ask what is the point of ideal theory and to show the nature of its value. I will also argue that, while the debate over the point of ideal theory can be productive, it will only be so if we avoid treating ideal and nonideal theories as rival approaches to political theory.

My argument will unfold as follows. I begin in section 1 by sketching the background of the debate over the distinction between ideal and nonideal theory. In doing so I will concentrate on two recent critiques of ideal theory developed by Colin Farrelly and Charles W. Mills, hoping to show that they miss their target.¹ I will then, in section 2, suggest why we may want to define ideal theory as theory that fails to issue recommendations for how to improve our society that are applicable for us here and now. I will argue there that while ideal and nonideal theory can be profitably distinguished, many, if not all, theories of justice, such as Rawls's theory of domestic justice, operate simultaneously at the ideal and nonideal level. In section 3 I will review and reject an alternative suggestion for distinguishing between ideal and nonideal theory and defend ideal theory from the criticism that good normative theory should eschew relying on false assumptions. Finally, in section 4, I will explain the point of ideal theory that does not issue recommendations for how to improve our society that are applicable for us here and now.

¹Colin Farrelly, "Justice in Ideal Theory: A Refutation," *Political Studies* 55 (2007): 844-64; Charles W. Mills, "'Ideal Theory' as Ideology," *Hypatia* 20, no. 3 (2005): 165-84.

1. The Ideal/Nonideal Distinction

What is ideal theory in the context of normative political theory? One straightforward understanding of the term states that a theory is ideal if it helps itself to the assumption of so-called full compliance.² Such a theory assumes that people, at least generally speaking, do what the theory demands of them. Such a theory is ideal in the sense that it assumes an idealized (untrue) version of reality: in the actual (nonideal) world, people do not always comply with what is required of them. The problem with such a straightforward definition of ideal theory is that even if assuming full compliance is a sufficient condition for a theory to qualify as an ideal theory, it is unclear if it can serve as a necessary condition. After all, there is no shortage of other features of the actual (nonideal) world that could potentially be bracketed off. What if we were to bracket off the fact that in the nonideal world people need access to food and water to survive? Would theories based on the assumption that people can live on air have enough in common with theories assuming full compliance to allow us to make common statements about their usefulness?

As recent work on ideal theory demonstrates, a number of theorists have tried to move beyond the understanding of ideal theory as based solely, or primarily, on the assumption of full compliance. Farrelly, for example, in a recent contribution to the debate, suggests (or implies) that all of the following might bear on whether a theory ought to be classified as ideal: (inappropriate) fact insensitivity; taking into account abstract hypotheticals; a cost-blind approach; ignorance of constraints; idealization; and ignorance of what is feasible.³ This is a long list of features and Farrelly does not explore in any detail how they are all related to one another. Ultimately, however, he suggests that we see (inappropriate) insensitivity to facts as the main defining characteristic of ideal theory.⁴ Specifically, the type of fact insensitivity that lies at the heart of ideal theory, according to Farrelly, involves idealization.⁵

²John Rawls, *A Theory of Justice*, revised ed. (Cambridge, Mass.: Harvard University Press, 1999), pp. 7-8 and 212; Michael Phillips, "Reflections on the Transition from Ideal to Non-Ideal Theory," *Noûs* 19 (1985): 551-70, pp. 553-56; Liam Murphy, "Institutions and the Demands of Justice," *Philosophy and Public Affairs* 27 (1998): 251-91, pp. 278-79.

³Farrelly, "Justice in Ideal Theory," pp. 844-48.

⁴The fact insensitivity must be (as I designated it) "inappropriate" because Farrelly accepts that some fact insensitivity is necessary in order to avoid normatively privileging the status quo. However, he adds that "[d]etermining exactly how fact-sensitive a theory of justice ought to be goes beyond the aspirations of this particular article" (*ibid.*, p. 846).

⁵*Ibid.*, p. 848. See Onora O'Neill, "Abstraction, Idealization and Ideology in Ethics," in J.D.G. Evans (ed.), *Moral Philosophy and Contemporary Problems* (Cambridge: Cambridge University Press, 1988), pp. 55-69, and her *Towards Justice and Virtue* (Cambridge: Cambridge University Press, 1996), esp. pp. 38-44.

Here Farrelly draws on a distinction between idealization and abstraction introduced by Onora O'Neill: *abstraction*, according to O'Neill, consists in ignoring some complexities of a given problem without, however, assuming any falsehoods about them; *idealization*, on the other hand, consists in theorizing that makes some false assumptions.⁶ Thus, for example, recommending that people be held responsible for their choices on the basis of the assumption that they can all choose wisely would involve idealization (since we are assuming a falsehood about a number of people), while recommending that people be held responsible for their choices because it often has positive incentive effects—whether or not they can choose wisely—would involve mere abstraction from the complex reality in which some can and some cannot choose wisely.⁷ It is precisely because ideal theory involves idealization, according to Farrelly, that it ends up being “impotent” as normative theory; on his view we must turn to nonideal theory if we want to make progress with normative questions.⁸

Farrelly's distinction between ideal and nonideal theory echoes that of Mills, who has also argued that ideal theory involves idealization. Mills, however, explicitly sets up the contrast between idealization and its lack in starker terms than O'Neill does herself. We engage in idealization, according to Mills, when we build a model of some P, which is not descriptive of what the P is like, but rather models, on the basis of assumptions that are *significantly* false, what an ideal P should be like (where “should” can refer to a host of values: moral, prudential, to do with efficiency, and so on).⁹ Ideal theory, then, according to Mills, stipulates *significantly* false attributes to individuals and/or groups and their interactions.¹⁰ The conclusion that Mills reaches about the use of ideal theory is similar to that reached by Farrelly: according to Mills, ideal theory cannot illuminate normative problems precisely because it involves assuming what is significantly false.¹¹

My outline of the views of both theorists has, of course, been only rudimentary (I will revisit their views in more detail in section 3 below), but the discussion should already have brought to the fore one fundamental question: how would we know if we arrived at a good definition of ideal theory and, by implication, a good account of the distinction between ideal and nonideal theory? Or, to put it differently: what is the dis-

⁶O'Neill, *Towards Justice and Virtue*, pp. 40–41.

⁷See, for example, Robert Goodin, “Political Ideals and Political Practice,” *British Journal of Political Science* 25 (1995): 37–56, pp. 40–45.

⁸Farrelly, “Justice in Ideal Theory,” esp. pp. 846–48 and *passim*.

⁹Mills, “‘Ideal Theory’ as Ideology,” pp. 167–68.

¹⁰*Ibid.*, p. 170.

¹¹*Ibid.*, p. 170 and *passim*.

inction supposed to capture? Neither Farrelly nor Mills in their accounts offers explicit answers to these questions. Farrelly does not consider objections to his construal of the category of ideal theory and it seems at times as if the category of ideal theory was constructed specifically in order to map onto the useful/useless distinction. Mills avoids the impression that he is simply setting out to map “ideal” onto “useless” by building his interpretation of what “ideal theory” is on the basis of an exploration of how we can interpret the word “ideal.”¹² But he too does not try to defend his categorization of ideal theory against alternative formulations. Admittedly, he suggests a hypothesis for how theorizing in the light of significantly false assumptions could attract enough theorists to emerge as a “type.” Such theorizing, he explains, serves the interests of those who happen to approximate the false assumptions made by the theory. Thus, for example, the assumption that all people relate to each other in an autonomous way, according to Mills, serves the interests of “middle-to-upper-class white males—who are hugely over-represented in the professional philosophical population.”¹³ So the ideal/nonideal theory distinction proposed by Mills serves the purpose of distinguishing those whose engagement in political theory is relatively uncritical from those who succeed in constructing theories that reflect more than their own interests.¹⁴

Disagreement over the distinction in question will, clearly, to some extent depend on the purposes that the distinction is supposed to serve. And, of course, the theorists who construct the ideal/nonideal labels might seek to further diverse ends with their labels. For example, the primary purpose of a theorist trying to distinguish ideal from nonideal theory may well be to reserve the label “ideal theory” for “useless theory” and then to try to identify and persuade us of what makes a theory useless. So in order to agree on a distinction, we will have to agree on the purpose(s) it is supposed to serve. To say more about the purposes of the distinction, however, let me first examine what this thing is whose typology is in question: normative theory.

¹²Ibid., pp. 166-70.

¹³Ibid., p. 172.

¹⁴In a recent book, Mills has returned to the problem of ideal versus nonideal theory. Regrettably, the book appeared too late for me to examine it in this article. But I should point out that Mills suggests in the book that part of the attraction of ideal theory derives, among other things, from its simplicity. I find this suggestion plausible. See Carole Pateman and Charles Mills, *Contract and Domination* (Cambridge: Polity Press, 2007), esp. pp. 103-4, 112-18, 232-34.

2. The Structure of Normative Theory

A theory—whether normative or positive—can be understood as a systematic account of our knowledge about a given dimension of reality (where the latter is broadly construed) that satisfies the criteria of what constitutes knowledge appropriate for that dimension. For example, a theory that systemizes our knowledge about climate fluctuations may need to meet criteria relating to falsification that would not apply to a theory dealing with the problem of theodicy. Importantly, all such theories have structures: that is, they have *inputs*, such as assumptions, *outputs*, such as final principles, and rules regulating the *derivation* of the output from the input. In other words, we can also think of theories as devices that are supposed to allow us to progress from one set of statements towards another set of statements in order to increase our knowledge about some X.

What is the internal structure of a *normative* theory?¹⁵ All normative theories must contain (normative) principles, that is, normative statements expressing position(s) on one or more values. This is what makes normative theories normative. In addition to principles, normative theories will need to contain elements that bind these principles together and account for their relevance. They will thus also contain models (however simple)—that is, analytic devices with a descriptive, explanatory, and/or predictive function—as well as arguments (statements) concerning the relationship between the principles. Following others, I am willing to apply the word “theory” generously to describe even a modest collection of coherent principles, of which some are the inputs and some are the outputs of a given theory. Some normative theories, however, will contain an additional type of output, which I will refer to as recommendations.¹⁶ Recommendations are specific proposals for actions, policies, and/or institutional reforms that are able to achieve improvements as measured by the specified principle(s).

Given that theories have structures, an obvious way of classifying them is in terms of the differences in their inputs, outputs, and/or rules of derivation. For example, the general distinction between normative and positive theories, which I invoked above, focuses on the presence or absence of (normative) principles among the inputs (and consequently

¹⁵In what follows I draw on an understanding of the structure of normative theories developed jointly with Alan Hamlin in Alan Hamlin and Zofia Stemplowska, “Theory, Ideal Theory, and the Theory of Ideals,” unpublished working paper.

¹⁶All normative theories, being normative, will of course contain (normative) principles and therefore they will, in some vague sense, all offer recommendations in the form of a defense of some principles. However, they will not all prescribe specific actions, policies, or institutional reforms that should succeed in advancing the values expressed by the principles.

also outputs) of a given theory.¹⁷ Similarly, the traditional distinction between ideal and nonideal theory focuses on the presence or absence of a certain input: the assumption of full compliance. Farrelly's and Mills's suggestions also lead us to focus on the input: the presence of false or significantly false assumptions. It seems to me that (outside of debates about disciplinary boundaries, which is not my focus here), 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. This is what in effect Mills and Farrelly are trying to do; they want to show that making false or significantly false assumptions will make a theory useless. In the next section I will argue that they fail. But, more importantly for now, my point is that the main purpose of drawing the distinction between ideal and nonideal theory is to differentiate normative theory with regard to its function (even if sometimes this means identifying a category of theories that serve no useful function at all). Having said that, focusing on outputs (and thus functions) of a theory in order to categorize it will often be just the first step towards answering the wider question of whether there are certain inputs that correspond to the stated functions. It is nonetheless a crucial step and if we want to categorize theories in accordance with their functions, we will in effect be inquiring about the type of output that normative theories produce.

One crucial difference between various normative theories concerns whether they offer viable recommendations, where by viable recommendations I mean recommendations that are both achievable and desirable, as far as we can judge, in the circumstances that we are currently facing, or are likely to face in the not too distant future.¹⁸ Let me call such recommendations "AD-recommendations" to emphasize that what they recommend is both achievable and desirable. In fact it is precisely the lack of AD-recommendations that is at stake in debates about the uses of ideal theory. Therefore, a straightforward way of drawing the ideal/nonideal theory distinction is to define nonideal theory as theory that issues AD-recommendations, and ideal theory as theory that does not.¹⁹ Importantly,

¹⁷Incidentally, different approaches to the rules of derivation (the difference in emphasis on the importance of cataloguing all the moves in the argument versus developing a distinct way of portraying a problem, often with the help of new phraseology) might be what distinguishes so-called analytical from so-called continental philosophy.

¹⁸I am grateful to Michael Otsuka for pressing me to clarify the temporal element in my definition of viable recommendation.

¹⁹Of course claiming that we should distinguish between theories that do and do not offer AD-recommendations does not make it any easier to spot AD-recommendations when we are searching for them or are confronted with them, and I will have to leave this issue unexamined. Indeed, this proposal is only preliminary, since for the purposes of my

doing so would immediately suggest why nonideal theory is seen as important (we all want to hear about AD-recommendations) and why ideal theory might have been subject to criticism (it does not offer AD-recommendations). At the same time, however, we would not be prejudging the answer to the question whether ideal theory is useful (must AD-recommendations be the only thing theorists care about?) and, most importantly, we will now be in a position to ask a question of some importance: what, if anything, is the purpose of constructing theories that do not issue AD-recommendations?

It is worth mentioning briefly how the distinction I have just proposed maps onto another possible distinction. Amartya Sen has recently suggested that we should distinguish between transcendental and comparative theories of justice.²⁰ The former are interested in theorizing about what ideal justice involves, while the latter are interested in ranking alternative societal arrangements, ultimately with the purpose of uncovering how to make advancements towards justice from where we find ourselves.²¹ Indeed, given that Sen sees the role of comparative theories as that of identifying ways of advancing justice,²² he is implicitly distinguishing not so much between transcendental and comparative theories of justice but between transcendental theories and a subset of comparative theories, those that rank *viable* societal arrangements. If this is correct, then Sen's categorization can be subsumed under the one I am suggesting: *to the extent that we are prepared to accept that ideal justice is not achievable for us here and now*, transcendental theories would fall into the ideal theory category while comparative theories would fall into the nonideal theory category.

Let me emphasize, however, that the distinction I am proposing is just one possible way of mapping the ground.²³ I am advocating it because it

argument here I have to side-step many difficulties such as, for example, under which category, if any, we should locate theories that aim to offer AD-recommendations but whose recommendations, as a matter of fact, turn out to be non-AD-recommendations (for reasons other than that the theories are simply empirically inaccurate). I am able to raise this issue only briefly in section 4, where I also discuss theories that aim to offer recommendations in the full knowledge that they will be non-AD.

²⁰Amartya Sen, "What Do We Want From a Theory of Justice?" *The Journal of Philosophy* 103 (2006): 215-38.

²¹This theme—the importance of distinguishing between theories that aim to reform what we already have and theories that do not recognize the status quo as a constraint—is also developed by Jonathan Wolff, "Harm and Hypocrisy: Have We Got It Wrong On Drugs?" *Public Policy Research* 14 (2007): 126-35.

²²Sen, "What Do We Want From a Theory of Justice?" pp. 236-38.

²³For further discussion of how to map the ground, see Ingrid Robeyns, "Ideal Theory in Theory and Practice," *Social Theory and Practice*, this issue, pp. 341-62; and Laura Valentini, "On the Apparent Paradox of Ideal Theory," *Journal of Political Philosophy*, forthcoming.

allows us to pose what I take to be an important question: what is the purpose of building theories that do not offer AD-recommendations? But one implication of the distinction, if it is adopted, is that most substantive theories dealing with complex matters such as justice in a reasonably exhaustive way will be neither entirely ideal nor entirely nonideal. Instead they will contain elements of both. That is, they will issue some AD-recommendations but they will also issue recommendations that are not AD-recommendations and will, in addition, discuss some problems without aiming to issue any recommendations about them at all. This does not mean that there is no point in drawing a distinction between ideal and nonideal theory—after all, it depends on what we want the distinction for—but it does mean that the distinction will be far less important than critics of ideal theory may have led us to believe. It also means that the distinction is not like that between, say, the theory of evolution and creationism, which, given that they are mutually exclusive, could not be combined in order to explain a given phenomenon. Rather, the distinction between ideal and nonideal theory is more like that between, say, rational choice theory and psychoanalytical theory: they can deal with different aspects of the same problem and, as such, can be combined should the need arise.

Having suggested how we can think about ideal theory, I will also want to defend the use of such theory to tackle normative questions. However, before I go on to outline what the point is of theories that offer no AD-recommendations, let me first explain why I do not think that there is much mileage in categorizing theories by whether they do (or do not) make false or significantly false assumptions.

3. What's the Point of False Assumptions?

I have argued above that the debate over ideal and nonideal theory should be understood as a debate over the function(s) of normative theory. One possible interpretation of the move to categorize theory in accordance with whether it makes false (and/or significantly false) assumptions is that such a move is intended to allow us distinguish between useful and useless, or informed and ill-informed normative theories. But, as I hope to show below, even if this is the intention of the ideal/nonideal theory categorization, it fails to deliver on its promise: making false or even significantly false assumptions does not condemn theory to be useless and may even be necessary for it to answer a range of questions.

To defend this claim, let me focus first on the uses of making false assumptions, leaving until later the problem of making significantly false assumptions (however we may want to define “significant” here).

Notice then that tackling broad problems (say, about what justice requires on a societal scale) inevitably requires making some false assumptions. For example, to theorize at a certain level of generality, we will have to make some approximations that, being approximations, will be false, but that will also be necessary to make any progress at all.²⁴ In addition to making approximations, theorists may have to pretend that certain difficulties do not arise or they will never be able to say what justice, on the whole, requires, as the complexity of the problem will be simply too great. If this is the case, however, then to define any theory that engages in idealization as ideal would be to confine the category of non-ideal theory to one concerned only with very narrow and relatively simple problems.

It could be objected that no idealization (understood as the making of false assumptions) is acceptable and that the category of nonideal theory can be used to capture this truth. Problems that cannot be tackled in their full complexity, the objection goes, should not be tackled at all: generality and incompleteness at the expense of the specific and the concrete is not a price worth paying for gaining an overall (but necessarily skewed) sense of where we stand vis-à-vis a given problem. The plausibility of this view clearly depends on a certain picture of meta-ethical reality that denies the validity of normative generalizations. I cannot here enter the complex debate about the merits of such a meta-ethical position; I can only point out that if accepting that idealization is useless requires one to accept such a meta-ethical position, then the price of purging idealization from nonideal theory is very high indeed. Such a theory will be unable to generalize about the problems it is supposed to solve. For the rest of the paper I will simply assume that this is not a price worth paying and that therefore we should neither reject idealization as a plausible move within theories, nor accept idealization as a sensible way of distinguishing between ideal and nonideal theory.

But what are we to make of theories that to reach their conclusions rely on *significantly* false assumptions? Of course the label “significant” is rather imprecise, but just as long as it is not meant to stand for “assumptions whose presence renders a theory useless” it can be left under-explored. What, then, can we learn that is of use by assuming the fanciful in our normative theories? First, by assuming the fanciful we can see more clearly just how crucial certain constraints are to shaping what we consider desirable or just when they are present (as long as we are careful not to assume away altogether the problem we are trying to

²⁴This incidentally explains why sometimes assuming full compliance—when full compliance approximates the truth—should not be viewed with suspicion even by those who believe that normative theory must offer useful prescriptions for us here and now.

tackle). Thus, for example, assuming that human nature was more malleable than we think it is would allow us to see how our conception of a relatively rigid human nature shapes what we consider to be just. This matters not merely for intellectual reasons but also because our knowledge of constraints—such as human nature—is subject to progress, and realizing the normative implications of the presence of a given constraint should affect our willingness to revisit and update our knowledge of that constraint.

Second, more specifically, we might make fanciful assumptions because we want to tease out the relationship between two principles (or, more broadly, the values that are captured by the principles) and it is easiest to do this in a radically simplified universe. This can best be illustrated by reviewing Farrelly's critique of Dworkin's account of liberal egalitarianism, which Farrelly attacks for falsely assuming that the full extent of human misfortune is narrower than it really is. In his theory, Dworkin is concerned with the question of what equality requires in a world in which people have differential talents and, sometimes, disabilities.²⁵ Why not, asks Farrelly, recognize that there are other misfortunes as well? He goes on to offer a list of other misfortunes that should be taken into account: crime, poor food preparation and sanitation, terrorist attacks, natural disasters, injuries at work, pollution, and so on.²⁶ The complaint here is that if we focus on just the misfortunes considered by Dworkin (and possibly other misfortunes, such as infertility, considered by Dworkinians),²⁷ we will end up overcompensating for these misfortunes and thereby not being able to compensate fairly for the other ones.

In response to this point, it is worth stressing first that political theory, like all scholarship, is useful only if it is cumulative, at least over some debates. I therefore see no reason why the fact that Dworkin focuses on a limited number of misfortunes renders his approach inapplicable to those who would like to see this list extended. But my main point is this: if we concentrate too much on assessing Dworkin's selection of misfortunes, we are in danger of missing the most innovative and fundamental insight of Dworkin's theory. This insight does not consist in the recognition that disabilities and lack of talent pose problems for egalitarians that other misfortunes do not, but in his point about what we should aim at when we are working out what is owed to people on account of inequalities

²⁵Ronald Dworkin, *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2000), esp. chap. 2.

²⁶Farrelly, "Justice in Ideal Theory," p. 857.

²⁷Farrelly references (p. 858) Justine Burley, "The Price of Eggs: Who Should Bear the Cost of Fertility Treatments?" in John Harris and Søren Holm (eds.), *The Future of Human Reproduction* (Oxford: Oxford University Press, 2000).

outside of their control. What Dworkin urges us to recognize is that we should not aim at ensuring that no one is worse off than others on account of inequalities outside of their control, but that, as far as possible, people should be compensated for such inequalities *to the extent of* what they would be willing to give up in order not to suffer them. If we do that, we will be able to determine, according to Dworkin, the point at which the costs of compensating for inequalities outweigh the costs of the sacrifices such compensation necessitates. In other words, however incomplete Dworkin's list of human misfortunes, he is still able to suggest how best to approach the problem of balancing our fundamental commitment, on the one hand, to make how people fare in life reflect their choices and ambition, and, on the other hand, to make how people fare in life immune from misfortune.

Claiming, as I am here, that even the most fanciful assumptions can be valuable is, admittedly, controversial. It means that it might be valuable to assume (to use Parfit's vocabulary) not only what is technically impossible—that is, what is practically impossible—but also what is deeply impossible—that is, what would require “a major change in the laws of nature, including the laws of human nature.”²⁸ However—as, for example, the debates about free will or personal identity show—assuming even the deeply impossible might be necessary to understand what we value. I am not thereby arguing that all problems that feed on fanciful assumptions are really worth exploring—only that some clearly are. I suspect that the critics of ideal theory would readily accept this conclusion and would clarify their position as one in which what is objected to is not the theorizing of the type I described above but only theorizing that does not issue AD-recommendations—often precisely because of the fanciful assumptions it is too ready to make. This, finally, brings me back to ideal theory as I defined it in section 2 above. Let me, therefore, in the final section tease out the uses of such theory.

4. What's the Point of Ideal Theory?

What are we then to make of theories that, at least as far as we can see, do not issue AD-recommendations? This is a question worth asking; after all, issuing AD-recommendations has been identified by some as the only purpose of acceptable normative theory. As John Dunn has put it, “[t]he purpose of political theory is to diagnose practical predicaments and to show us how best to confront them.”²⁹ As Farrelly further argued,

²⁸Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), p. 388. I am grateful to Jonathan Quong for drawing my attention to this distinction.

²⁹John Dunn, *Interpreting Political Responsibility* (Cambridge: Polity Press, 1990), p.

if trying to implement "... the conclusions of a theory [without adjustments] would not result in any noticeable increase in the justness of one's society, then it fails as a *normative* theory."³⁰ And according to Liam Murphy, "[i]f our theory has implausible implications for the non-ideal case ... it would fail as a normative political theory."³¹ Must normative theory guide us through issuing AD-recommendations (or at least imply such recommendations) in virtue of its aspiration to be normative?³²

Notice that normative theories may fail to issue AD-recommendations for different reasons (difficult as it may be to pull them apart when they are in operation): (i) they can be theories that offer recommendations that are not AD-recommendations, or (ii) they can be theories that do not aim at issuing any recommendations. The main purpose of theories that do not aim at issuing any recommendations seems to be simply to gain a clearer understanding of our values and principles. I find it hard to imagine that anyone might object to such theorizing or that there is much mileage in debating whether this type of analysis should merit the name of normative theory.³³ Let me therefore focus on the category of theories

193, quoted in Andrew Mason, "Just Constraints," *British Journal of Political Science* 34 (2004): 251-68, p. 253.

³⁰Farrelly, "Justice in Ideal Theory," p. 845 (emphasis in original). The whole quote reads: "If the collective aspiration to implement the conclusions of a theory would not result in any noticeable increase in the justness of one's society, then it fails as a *normative* theory." This, as far as I can tell, is an unfortunate phrasing of what Farrelly has in mind, since it suggests what Farrelly critiques, namely, that we should ask what justice requires while assuming that there is full compliance (i.e., a collective aspiration to implement the conclusions of a theory). But Farrelly is critical of the assumption of full compliance: "by assuming full compliance, ideal theorists violate the constraints of a *realistic* utopia" and side-step questions that they ought to address (*ibid.*, p. 845, emphasis in original).

³¹Murphy, "Institutions and the Demands of Justice," p. 279. Depending on whether something can be said to be implied by a theory even when this is not the authorial intention, there are two readings of which set of theories this principle is meant to criticize.

³²Farrelly and Dunn adopt a rather narrow understanding of what it means for a normative theory "to guide." This point is best illustrated not by a theorist but by a practitioner of real politics: "if we are to work for progress, we have to set a target we have not yet attained; if we are to keep aspiring to progress incessantly we have to aim at the impossible that can never be attained," wrote the Polish prince Adam Czartoryski in 1803 to the Tsar of Russia. Adam Jerzy Czartoryski, *Pamiętniki i Memoriały Polityczne, 1776–1809*, ed. J. Skowronek (Warszawa, 1986), p. 509, quoted (in Polish) in Andrzej Nowak, "Europa Narodów—Wizja Księcia Adama Jerzego Czartoryskiego," *Polski Przegląd Dyplomatyczny* 1 (2001): 179-206, p. 206. For an interesting suggestion about how normative theory can guide without offering any precise policy recommendation, see John Rawls, *Lectures on the History of Political Philosophy*, ed. Samuel Freeman (Cambridge, Mass.: Harvard University Press, 2007), pp. 3-9. But I do not want to rest my defense of ideal theory on a point about the meaning of "to guide."

³³In his article in this issue, Adam Swift examines why gaining a clearer understand-

that do issue recommendations, but they turn out not to be AD-recommendations.

What are we to make of theories that do issue recommendations but those that are inapplicable for us here and now? This happens when a theory does not build in (ignores) certain constraints (sometimes by building in other, inaccurate constraints) and this affects the type of solution that it delivers. Putting aside, for now, cases in which we are simply dealing with bad theory, we can distinguish between two further types of theories contained within the set: (a) theories that fail to issue AD-recommendations because they ignore the fact of nonmarginal noncompliance, and (b) theories that fail to issue AD-recommendations because, even with full compliance, there is no solution to the problem for which recommendations are sought. I will, in section 4.1 below, defend the point of theories that assume unrealistic levels of compliance, before, in section 4.2, taking up the harder task of defending the point of theories that search for solutions to problems that currently have no solutions that can be stated as AD-recommendations. I will then, in section 4.3, return to theories that fail to issue AD-recommendations because they misrepresent the constraints on what is desirable and achievable for us here and now.

4.1. Ignoring noncompliance

Imagine a theory that designs its recommendations while assuming that people will comply with all that can be justifiably asked of them, when, in fact, there is significant noncompliance (whether or not it is widespread or isolated). Think here, for example, of a theory concerned with recommending how to efficiently and justly solve the problem of poverty. Imagine that such a theory recommends that all affluent people donate a small amount of their wealth to a global fund. Given that many people would fail to do so, such a theory would not succeed in offering a solution to the problem of poverty that is achievable for us here and now. There is a sense, of course, in which the solution, even if never achieved, is “achievable,” since it could be achieved if people chose to act on it. But it is a sense that has proven too thin for many and for this reason such theories can be classified as theories that fail to offer AD-recommendations.

Clearly, then, the first major worry regarding the assumption of full compliance (or any inadequate level of compliance) is that a theory that makes the assumption will be unable to offer any specific guidance re-

ing of our values and principles would help us to come up with better AD-recommendations. Adam Swift, “The Value of Philosophy in Nonideal Circumstances,” *Social Theory and Practice*, this issue, pp. 363-87.

garding what to do in circumstances of noncompliance. After all, actions and institutions whose desirability is assessed in the light of the assumption of full compliance will often be undesirable given the actual extent of noncompliance. This is because the level of compliance affects the costs and benefits of actions and institutions: so, failing a student for a bad essay, when other teachers also follow the rule to fail students who submit bad essays, has a different moral significance from failing a student for a bad essay when no other teacher follows this rule. The only sensible response here is to say that in addition to theorizing about what justice requires under the assumption of full compliance, we will need to ask what it requires given the actual extent of noncompliance. Why, then, a skeptic may ask, should we bother with assuming full compliance to begin with?

There are two reasons why we must sometimes make the assumption nonetheless. First, unless we know what is desirable when there is full compliance, we could adopt a direction of reform for nonideal circumstances that unnecessarily moves us away from the ultimate aim of full compliance. For example, even if we think it is often justified for parents to send their children to private schools when other parents do so, it should make a big difference to the education policies we advocate if we also establish that in circumstances of full compliance there should be no private education.³⁴ Working out what is ideal under the assumption of full compliance forces us to inquire if we could innovate in order to come closer to what such a theory recommends and if we really *must* depart from the ideal when such departures are on the table. Assuming full compliance therefore has its straightforwardly practical role.

In addition, normative theorists should never dismiss the importance of knowing what a fully compliant society would be like: one of the roles of normative theory is to recommend more or less straightforwardly achievable change in the circumstances we find ourselves in, but another is to judge what we have already achieved against a final landmark of where we ought to be. Indeed, being concerned with the latter is part and parcel of accepting that there are good reasons for a given society to change in the first place. This is because advocating change requires accepting both that people, individually or collectively, can act better than they do and that they do not always act as they should (hence the call for change). But if people do not always act as they should, then the assumption of full compliance guards us from making our demands too permissive towards those who are simply unwilling to comply with reasonable

³⁴For an argument along these lines, see Adam Swift, *How Not to be a Hypocrite: School Choice for the Morally Perplexed Parent* (London: Routledge, 2003).

limits on the permissibility of their actions. What motivates the assumption of full compliance, then, is the worry that without it the normative requirements we come to acknowledge will be normatively inaccurate. The assumption of full compliance guards us from accepting, as David Estlund has put it, the implausible dictum that “ought implies reasonably likely.”³⁵

Nonetheless, the critics of the assumption may try to mount an objection to the assumption, suggesting that the costs of making it are simply too great. Assuming full compliance, they can argue, may result in a theory of justice that actually wrongly demands compliance where individuals should instead be entitled to choose what to do.³⁶ Thus some theorists worry that people’s persistent unwillingness to do F may be an indication that their legitimate interest is best served by not demanding that they do F.³⁷ For example, people’s unwillingness to donate money to relieve disasters abroad could be (mis)read as an indication of their legitimate interest in leading a life that is largely free of positive demands on their resources. Furthermore, some have taken people’s persistent unwillingness to do F as an indication that the requirement to do F is somehow too demanding and thereby genuinely impossible to fulfill, which disqualifies it as a demand of normative theory given the dictum that ought implies can. For example, if people are not as attentive to others as they are asked to be by moral theorists, it might indicate that they simply do not have the mental capacity to deal with all that the moral theorists have asked them to deal with.³⁸

These are all genuine difficulties, but what they suggest is not that we abstain from making demands on people but that we are careful in formulating them.³⁹ After all, abstaining from making demands is not a

³⁵David M. Estlund, “Utopophobia: Concession and Aspiration in Democratic Theory,” in *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press, 2007), pp. 258–75, at p. 265. Unfortunately, Estlund’s book only appeared after this paper was accepted for publication and thus I cannot draw as much as I would like to on his defense of what he calls hopeless (in the sense of unrealistic) theories.

³⁶Compare here Thomas Nagel, “The Problem of Utopianism,” in *Equality and Partiality* (Oxford: Oxford University Press, 1991), and Mason, “Just Constraints,” pp. 260–62.

³⁷For some of the most illuminating discussions of this problem, see Liam Murphy, “Over-Demandingness, Alienation, and Confinement,” in *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press, 2000), pp. 9–33; and Peter Unger, “Living High and Letting Die Reconsidered: On the Costs of a Morally Decent Life,” in *Living High and Letting Die* (Oxford: Oxford University Press, 1996), pp. 133–57.

³⁸Demands can be demanding, of course, without being impossible to fulfill, but if this is the case then we are back to the preceding problem of how to distinguish between a mere unwillingness to follow a demand and an unwillingness indicative of an entitlement not to follow the demand.

³⁹Interestingly, this may have implications for the precision of such demands. As

morally neutral position we can fall back on. No normative theory can side-step the difficulty that arises because we are not sure how to interpret people's unwillingness to comply by deciding not to ask them to comply with more than they are themselves willing to. All in all, therefore, making the assumption of full compliance, even when it is strikingly false, and despite the genuine difficulties that come with it, is necessary to ensure that we know what a commitment to a given set of values entails. We need to know this partly so that we know what to strive for and partly so that we know how we can judge each other. This does not mean that we do not need theories that lift the assumption—we do—but it does mean that we also need theories that stick with it and tell us how the world ought to be.

4.2. *Ignoring feasibility*

What, however, are we to make of theories whose recommendations would remain non-AD-recommendations, even if people complied with everything that could be legitimately required of them but, for example, the historical circumstances or the level of economic development would still ensure that the recommendations were not viable? Possibly the most famous example of such a theory comes from Marx. That is, for at least some Marxists, Marx's theory should be read precisely as offering recommendations about how to achieve the spontaneous harmony of people's aims. One such recommendation, for example, is to abolish the private ownership of the means of production. This is not the place to examine the finer details of this recommendation, so let me for the sake of the argument just state that, in my view, this is hardly an AD-recommendation.

What are we to make of this recommendation, then? One response would be simply to conclude that this is a pretty bad recommendation: in our world we cannot reasonably demand that people accept the abolition of the private ownership of the means of production. But there is also a more sophisticated reading of the function of this recommendation. According to this reading, the recommendation shows that *if we want* a certain type of society—a communist society—then, among other things, we must abolish private ownership of the means of production. This does not mean that we should *actually* try to do it here and now. Rather, it means that this is what it would take to have a communist society. The theory simply models “the conditions under which the spontaneous coherence of the aims and wants of individuals

noted by Leif Wenar, imprecise demands on the affluent to help the global poor may cause frustration and thus “spur avoidance of the entire topic.” Leif Wenar, “What We Owe to Distant Others,” *Politics, Philosophy & Economics* 2 (2003): 283–304, p. 297.

is neither coerced nor contrived but expresses a proper harmony consistent with the ideal good.”⁴⁰ The recommendation to abolish the private ownership of the means of production is part of a set of conditions that tell us what it would take to achieve some outcome, in this case a proper harmony of human ends.

This raises two questions. First, we might wonder what is the normative purpose of issuing recommendations that we should not actually follow. Second, we might even resist the suggestion that non-AD-recommendations can still count as recommendations. Let me begin with the latter question in order to defend my suggestion that it makes sense to talk about non-AD-recommendations. First, note that there is nothing mysterious about achievable but undesirable recommendations. Indeed, recommendations issued by theories that ignore feasibility constraints could easily turn out to be undesirable once the actual costs of implementing them are taken into account.⁴¹ Moreover, there is nothing mysterious about unachievable recommendations if we qualify as unachievable (as I did in the previous section) proposals that fail to motivate people to comply with them even though they do not demand the impossible. That is, again, the idea of a recommendation that fails to motivate makes perfect sense. This brings us to a tougher question: are proposals advocating what is in fact unachievable for us, even if we were motivated to achieve it, really recommendations?⁴² The worry here seems to be that recommending what is unachievable for us in this sense offers no guide to action.⁴³ But even without going into the details of the debate over whether ought implies can, there is at least a sense in which conditional recommendations can be classified as recommendations. Thus some proposals might advocate an action (or policy or institutional set-up) that could be followed under the right circumstances but it so happens that the circumstances simply do not materialize, thus making it impossible to act in the recommended way. Think here, for example, of the recommendation (if it is, as I am suggesting, a recommendation) to help save a drowning child in a pond, even if the circumstances in which one could save a drowning child might never materialize.⁴⁴

⁴⁰Rawls, *A Theory of Justice*, p. 249.

⁴¹This is indeed why Farrelly objects to ideal theory.

⁴²I am grateful to an anonymous referee for pressing me to clarify my position on this issue.

⁴³But see again n. 32.

⁴⁴More tentatively, we might discover that the distinction between proposals that are not recommendations and proposals that are tracks the distinction between deep impossibility on the one hand and technical impossibility or possibility on the other. Imagine a policy proposal stating that doctors should assign certain medical treatments on the basis of information about brain activity during one's normal day (as measurable, in principle, by brain scanning technology such as fMRI). I think that we could classify such a pro-

However, even if I am wrong and proposals recommending the unachievable should not be referred to as recommendations, it would still not follow that such proposals have no normative function. This brings me to the second question raised above: what is the purpose of theories issuing non-AD-recommendations and/or pseudo-recommendations? One answer, already mentioned in the previous section, is that the purpose of normative theory is not only to guide action but also to evaluate states of affairs.⁴⁵ But there is another, more surprising, reason as well: sometimes we need to ask what realizing certain values or principles requires in the absence of (some of) the constraints that we are currently facing—even if it means issuing only non-AD-recommendations—in order to identify more of what realizing this value may require of us here and now. By saying this I do not want to deny that it should be obvious what acting on a given value requires across many dimensions even if we do not ask ourselves what realizing this value might require in the absence of some existing constraints. Thus, for example, it should be obvious that eliminating easily avoidable suffering and giving the accused fair trials advances justice. That is, to adopt these two aims for ourselves we do not need to examine what justice might require in the absence of some existing constraints on it (call this proper justice), and even less so what justice might require in the absence of all constraints that might prevent us from reaching an ideally just society (call this ideal justice). As Sen has suggested, to imagine that we always need to know what ideal justice requires in order to know what would advance justice for us here and now would be like insisting that we need to know that Mount Everest is the tallest mountain in order to compare the size of two smaller ones.⁴⁶

Nonetheless, reflection on proper or even ideal justice is essential to help us identify all that justice does require of us here and now. This is because justice, unlike height, is a complex value.⁴⁷ Let me explain the

posal as a recommendation even if it is technically impossible to use the brain scanning technology to measure brain activity in everyday situations. The proposal identifies a coherent policy whose fulfillment conditions we can understand. On the other hand, suggesting that we turn frogs into princes would not qualify as a recommendation since, at least at present, we must suspect it to be deeply impossible. I am grateful to Jonathan Quong for discussion of the relevance of Parfit's distinction to the purposes of normative theory.

⁴⁵Swift, "The Value of Philosophy in Nonideal Circumstances," pp. 366-68; G.A. Cohen, "Facts and Principles," *Philosophy and Public Affairs* 31 (2003): 211-45, p. 243.

⁴⁶Sen, "What Do We Want From a Theory of Justice?" p. 222.

⁴⁷One way of understanding the difference in complexity between height and justice is that for any discernible change in an object whose height is at stake we can straightforwardly tell if a given change makes the object higher or lower but the same is not true about change in society with respect to justice.

point of reflection on proper or ideal justice by analogy with relationships. Imagine, then, that you are searching for a life partner. In such cases it helps to reflect on what an ideal partner would be like even if you do not know if one will be available. This is not because without such reflection we will not know that we ought to avoid psychopaths. Rather, it is because such reflection, when done properly, helps us understand which seeming vices might be virtues (even an ideal partner would have them), and which vices, even if unavoidable among the available candidates, are truly regrettable. Moreover, it is only if we are confident that we are reflecting upon an ideal (or at least undeniably good) partner—rather than simply on the available candidates—that we can accept such revelations about vices and virtues as insights into good relationships rather than available relationships. In essence, my argument is that existing constraints on what we can achieve may distort our understanding of what a given value requires and that reflection on the value when such potentially distorting constraints are absent should help us spot such distortions. If so, then comparing what we aim to achieve here and now (when what we can achieve here and now is messy and limited) with a less constrained ideal allows us to test our grasp of values against situations in which our judgments are less fallible.⁴⁸

Looking at the work of Rawls and Marx can illustrate this in relation to the value of justice or societal harmony. After all, one way of interpreting Rawls's project is to see it precisely as an attempt to model the conditions under which *justice*—as opposed to a *modus vivendi*—is possible.⁴⁹ Rawls, of course, may be most interested in this question *to the extent* that the conditions turn out to be realistic under reasonably favourable conditions, but even if he is wrong about what is within our reach, and thus fails to offer us AD-recommendations, the immense value of his project is that he asks *what at a minimum is required to achieve undistorted justice*.⁵⁰ To interpret Rawls's project this way is to

⁴⁸Compare Phillips, who notes that “[n]onetheless, there does seem to be some important connection between what we take to be true ideals and what we regard as duties under various historical conditions ... historical codes may be paired with ideal codes in virtue of instantiating the same fundamental value or imperative.” Phillips, “Reflections on the Transition from Ideal to Non-Ideal Theory,” p. 564. What he does not add but what I am emphasizing here is that we need the ideal codes in order to make sure that the historical codes do indeed capture what we value.

⁴⁹Rawlsian ideal theory defines the conditions under which individuals can have “a reasonably harmonious and stable pluralist and democratic society.” John Rawls, *Justice as Fairness* (Cambridge, Mass.: Harvard University Press, 2001), p. 196.

⁵⁰Incidentally, I believe that Rawls asks this question even if G.A. Cohen is right to point out that there is a difference between fundamental principles of justice and principles of regulation and even if Rawls were to misinterpret what justice fundamentally requires. See G.A. Cohen, *Rescuing Justice and Equality* (Cambridge, Mass.: Harvard University Press, 2008).

notice that he is not necessarily answering the question of “how we can get to justice from where we are right now,” but rather the question of “how far the ideal of justice can be rendered compatible with constraints that limit the extent to which it can be realized before it is changed beyond recognition” (with the hope that it will reveal justice to be compatible with the constraints currently faced by liberal democracies).⁵¹ It is precisely because we can be reasonably certain that we are inquiring about justice—rather than simply about what our political system can currently deliver—that, if we accept Rawls’s argument, we will see that justice does not require, for instance, that we aim to eliminate disagreement over conceptions of the good. Similarly, those who do value communist ideals may never learn from Marx how to bring communism about but may, nonetheless, be persuaded that, say, emancipation is more valuable than liberal (negative) freedom. Even if they never tried to bring communism about, they could decide that reducing alienation is an important task for us here and now.

Reflection on ideal or proper justice on its own cannot tell us which actions, policies, and institutions to adopt, but it can, as I have argued, alert us to the presence of previously undiagnosed vices and virtues or expose some apparent vices as virtues. This is important since there is no good reason to expect that all of the requirements of justice should present themselves to us as obvious just as soon as we correctly identify all of our constraints. All in all, therefore, constructing theories with non-AD-recommendations need not be merely an exercise in cleverness: issuing recommendations that are not AD-recommendations is a task for normative theory when it allows us to understand what our values require of us in the situation in which we find ourselves.

4.3. Misrepresentation of what is desirable and achievable

What then of theories that offer non-AD-recommendations for reasons other than that these theories help us to identify which goals we should set ourselves and how we should evaluate where we find ourselves? Such theories, I want to claim, are not theories that we must defend in order to defend ideal theory. Rather, we should recognize them as failed normative theories *tout court*. If all theorizing were properly done, this final category of theories should simply be an empty set. But I hope to have shown above that Mills is wrong when he suggests that in cases in which actual circumstances are recognized to depart from what the theory as-

⁵¹That this is the correct reading of Rawls’s project is further illustrated by Rawls’s discussion of the transition from *A Theory of Justice* to *Political Liberalism*. John Rawls, *Political Liberalism*, paperback ed. (New York: Columbia University Press, 1993), pp. xvi-xvii.

sumed, ideal theory “claims that starting from the ideal is at least the best way of realizing it.”⁵² Good ideal theory can serve useful purposes even if it cannot offer AD-recommendations, and it therefore does not have to claim that it offers a starting point for change. Of course, many theories can both say important things about our values as well as what we can justifiably expect of people, and at the same time wrongly claim that their recommendations are AD-recommendations when they are not. It is precisely because such a combination is possible, and because ideal theory is necessary to normative inquiry, that political theorists will inevitably keep constructing theories that do some things well and some badly. The point then is to know how to interpret them well.

5. Conclusion

I have argued in this paper that one helpful way of understanding the cut between ideal and nonideal theory is roughly to classify as nonideal any theory that issues recommendations that are desirable and achievable for us given where we find ourselves and, as ideal theory, any other type of normative theory. Ideal theory, if properly executed, not only has a point: it is indispensable to help us make progress with a number of key questions. Far from being irrelevant or misleading, it helps us to see our principles and problems more clearly, it ensures that even when we are not motivated to do what can be required of us we are not thereby let off the hook, and it allows us to uncover more of what we value and should therefore pursue. Ideal theory therefore uncovers, clarifies, and safeguards our normative commitments.

Once we understand the cut between ideal and nonideal theory the way I have suggested in this paper, we will notice that complex normative theories, such as Rawls’s theory of justice, are likely to contain within themselves both ideal and nonideal theory. This suggests that there is no real conflict between ideal and nonideal theory. Indeed, we should see all normative theorizing as part of a common project: the use of ideal theory techniques contributes to the use of nonideal theory techniques (and vice versa). Of course, it could be argued that political theorists have already made extensive use of ideal theory while neglecting nonideal theory. This may well be the case. However, questioning the purpose of ideal theory as such will simply take us a step backwards

⁵²Mills, “‘Ideal Theory’ as Ideology,” p. 168. Note that Rawls, for instance, is explicit about the limits of the direct applicability of his theory: “In the more extreme and tangled instances of nonideal theory this priority of rules [the lexical priority of the liberty principle over the other principles of Rawlsian justice] will no doubt fail; and indeed, we may be able to find no satisfactory answer at all.” Rawls, *A Theory of Justice*, p. 267.

rather than help us with what we should be striving to achieve, namely, bringing political science and normative theory closer together if we are to solve the urgent practical questions that confront our societies.⁵³

Zofia Stemplowska

Manchester Centre for Political Theory
Politics, School of Social Sciences
University of Manchester
zofia.stemplowska@manchester.ac.uk

⁵³I am grateful to Alan Hamlin, Ben Jackson, Patti Tamara Lenard, Daniel McDermott, Steven Montgomery, Avia Pasternak, Jonathan Quong, and Laura Valentini, as well as the editors of this special issue, Ingrid Robeyns and Adam Swift, and an anonymous referee for comments on an earlier version of this paper, and to Sorin Baiasu, Katherine Eddy, Mark Philp, Julia Skorupska, and Hillel Steiner for helpful discussion. I have also benefited from the comments of the participants at the MANCEPT internal political theory seminar on 14 March 2007; the 2007 ECPR Joint Sessions workshop “Social Justice: Ideal Theory, Non-Ideal Circumstances”; the Autumn 2007 UCL Priority in Practice Conference and the Nuffield Political Theory Workshop. Much of the argument in this paper draws on joint work with Alan Hamlin in “Theory, Ideal Theory, and the Theory of Ideals.”