

## Ideal Theory in Theory and Practice

### 1. Introduction

In the post-Rawlsian literature on theories of justice, most of the work done by mainstream political theorists and philosophers is part of what is known as “ideal theory.” John Rawls explicitly defined his work as ideal theory, which he described as a conception of a fully just society. He acknowledged that, with the exception of his analysis of civil disobedience, he would not pursue nonideal theory. Rawls thought he prioritized ideal theory for good reasons: “the reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems [that we are faced with in everyday life].” He went on to claim that “the nature and aims of a perfectly just society is the fundamental part of the theory of justice.”<sup>1</sup> Rawls believed that we cannot develop nonideal theory without first working out ideal theory. Post-Rawlsian theories of justice have broadly endorsed this view about the need for theorists of justice to prioritize ideal theory, and the current literature on social justice is mainly concerned with working out, revising, refining, and debating the relative merits of different ideal principles of justice, and the justification of those principles.

In recent years, however, a growing number of political philosophers have expressed worries about the nature of ideal theory and its dominance in the literature on social justice. Jonathan Wolff has argued that “ideal thinkers who want to have some impact on reality should pay more attention to issues of transition.” Amartya Sen considers ideal theory (which he calls “transcendental theory”) neither necessary nor sufficient to guide justice-enhancing policies. He strongly criticizes what he takes to be the “all or nothing extremism” of ideal theory, and argues for moving the theory of justice outside that “little corner.” Charles Mills contends that ideal theory “is really an *ideology*, a distortional complex of ideas, values, norms, and beliefs that reflect the nonrepresentative in-

---

<sup>1</sup>Rawls, *A Theory of Justice*, revised ed. (Cambridge, Mass.: Harvard University Press, 1999), p. 8.

terests and experiences of a small minority of the national population—middle-to-upper-class white males—who are hugely *over-represented* in the professional philosophical population.” Colin Farrelly has also staged a fierce attack on ideal theory, claiming it to be “inherently flawed.”<sup>2</sup>

What is one to make of these strong criticisms of ideal theory? What are the merits and limitations of ideal and nonideal theory, and what is their proper role? These are the broad questions that I will address in this essay.<sup>3</sup> Yet before we can embark on that task, we have to confront the problem of the rather different and often conflicting definitions that are given to “idealize,” “ideal theory,” and “nonideal theory.” I will therefore first clear the ground by proposing a set of definitions and descriptions, together with a simple typology of the different types of work that can be distinguished in the normative social justice literature. One of the key questions when defining ideal theory is the question of how it relates to idealizations. I will distinguish ideal theory from idealizations and discuss how they relate. I will suggest how we could distinguish useful idealizations from bad idealizations. The last part of the paper briefly discusses the implications for the practice of ideal theorizing. I will argue that the role of ideal theory is limited, perhaps more so than generally acknowledged in the literature, and that these limitations should be much more explicitly discussed by ideal theorists. I will defend the view that ideal theory does have a role to play, but that in the daily practice of theorizing about justice at least three changes are needed: more attention should be paid to filtering out bad idealizing assumptions, the profession should correct the current academic bias towards ideal theorizing by re-evaluating nonideal theory, and much more theoretical work remains to be done on the questions of how to bridge the gap between ideal and nonideal theory, and on clarifying what makes them good rather than bad theories.

## 2. Normative Social Justice Analysis: A Typology

What are the different types of normative social justice research that we can distinguish?<sup>4</sup> There are, in my view, three different layers of research

---

<sup>2</sup>Jonathan Wolff, “Fairness, Respect, and the Egalitarian Ethos,” *Philosophy and Public Affairs* 27 (1998): 97-122, p. 113; Amartya Sen, “What Do We Want From a Theory of Justice?” *The Journal of Philosophy* 103 (2006): 215-38, pp. 235, 238; Charles W. Mills, “‘Ideal Theory’ as Ideology,” *Hypatia* 20 (2005): 165-84, p. 172; Colin Farrelly, “Justice in Ideal Theory: A Refutation,” *Political Studies* 77 (2007): 844-64, p. 845.

<sup>3</sup>This essay will have a rather exploratory nature. I will propose some working definitions and highlight some distinctions that I think can advance our understanding of these questions. Nevertheless, the present arguments will remain limited to scratching the surface of a set of difficult metatheoretical questions.

<sup>4</sup>I am putting aside work on social justice that is primarily explanatory, descriptive,

that need to be distinguished: ideal theory, nonideal theory, and action design and implementation. The last two categories form the nonideal work in normative social justice research—but they are distinguished by the fact that action design and implementation is primarily empirical research, in contrast to nonideal theory.

In the following typology, I am not including the radically fact-insensitive kind of “pure” ideal theory that is best known from the work of G.A. Cohen.<sup>5</sup> This kind of theory does not start from a concern with addressing social ills and contributing to justice-enhancing practice, but rather with an interest in knowledge for the sake of knowledge. The corresponding conceptions of justice may therefore never be realizable, not now, not in the near future, and not even in any future that is consistent with established facts of the biophysical sciences, such as the fact of gravity or human mortality. Cohen is very clear about his view: “the question for political philosophy is not what we should do but what we should think, even when what we should think makes no practical difference.”<sup>6</sup> One could question whether such a “pure” view of the concept of justice should be considered *normative* at all. Such a view may have an inspirational and pure philosophical functioning, but one can raise serious doubts whether it has any relevance for the practice of justice beyond such an inspirational role. This kind of hyper-ideal fact-insensitive “pure” theory raises questions that are even more difficult than the questions raised by other types of ideal theory. An analysis of hyper-ideal fact-insensitive “pure” theory will therefore have to wait for another occasion, and we will ignore this category in the typology that follows.<sup>7</sup>

### 2.1. Ideal theory

The aim of ideal theory is to work out the principles of justice that should govern a society, that is, to propose and justify a set of principles of justice that should be met before we would consider a certain society just. In Rawls’s words, we ask “what a perfectly just society would be like.”<sup>8</sup> When defending and justifying the ideal principles of justice, we assume full compliance with those principles. However, the often heard shortcut

---

and non-normative, such as empirical research on what views about social justice are held by ordinary people. See, for example, David Miller, “Distributive Justice: What the People Think,” *Ethics* 102 (1992): 555-93.

<sup>5</sup>See G.A. Cohen, “Fact and Principles,” *Philosophy and Public Affairs* 31 (2003): 211-45.

<sup>6</sup>*Ibid.*, p. 243.

<sup>7</sup>See also Laura Valentini, “On the Apparent Paradox of Ideal Theory,” *The Journal of Political Philosophy*, forthcoming, section 1.

<sup>8</sup>Rawls, *A Theory of Justice*, p. 8.

among political philosophers that “ideal theory equals full compliance” is not very accurate and is potentially misleading, since full compliance may also hold for principles of justice that do not lead to a just society. When defining ideal theory, it should therefore be stressed that it is not about full compliance with any kind of principles of justice, but full compliance with those principles of justice that are morally required in order for society to be completely just.

Ideal theory can be *comprehensive* or *partial*. If an ideal theory of justice is truly comprehensive, then that theory would tell us what conditions should be met before each and every instance of injustice is removed. That is one extreme on a continuum where perhaps no contemporary theory is situated. Rather, some ideal theories are more comprehensive than others. Partial ideal theory can be partial in several ways. First, it may be partial if it specifies the minimal principles of justice, while leaving open the possibility that if these principles are met, further principles of justice would need to be achieved. An example is Martha Nussbaum’s capabilities approach, which argues that threshold levels of ten central human capabilities should be met as the first priority of justice, while leaving open what justice requires once these thresholds are achieved by all.<sup>9</sup> A second way in which ideal theory can be partial is by focusing on one domain of justice, such as justice in health, family justice, gender justice: we may defend principles of justice telling us what is required for complete gender justice, while remaining silent on all other domains of justice. Or ideal theory may be partial in a geopolitical sense, for example, by specifying the conditions that should be met for justice to be achieved within a nation-state, thereby disregarding justice between nation-states or on a global scale. Ideal theory may also be partial by restricting itself to spelling out the principles of political justice only, that is, describing what justice requires from the political institutions and agents, thereby disregarding what justice may require within private associations such as universities, religious organizations, or families. Partial ideal theory could also combine several of these partialities, for example, by specifying the principles required in order to achieve a threshold level of justice in health within the borders of one country.

What is the goal of ideal theory so defined? Ideal theory functions as a mythical *Paradise Island*. We have heard wonderful stories about Paradise Island, but no one has ever visited it, and some doubt that it truly exists. We have a few maps that tell us, roughly, where it should be situated, but since it is in the middle of the ocean, far away from all known societies, no one knows *precisely* where it is situated. Yet we dream of

---

<sup>9</sup>Martha Nussbaum, *Frontiers of Justice* (Cambridge, Mass.: Harvard University Press, 2006).

going there, and ask ourselves how we could get there, and in which direction we should be moving in order to eventually reach Paradise Island.

Paradise Island can serve as a metaphor for ideal theory. We don't know whether it can be reached and no one has ever set foot on the Island. Yet since it is our dream to go there, reaching Paradise Island is our ultimate goal. It gives us the direction in which we should be moving to reach a (minimally) just society, or a society that is just with respect to a particular domain. In other words, whether partial or comprehensive, ideal justice allows us to determine whether (partial) justice is achieved. Ideal theory specifies a number of conditions that have to be met before we consider a certain state of affairs as just. Take Rawls's theory of justice: it clearly spells out which principles of justice should be met before a society can, according to Rawls's view, be considered fully just. The same holds for partial theory. For example, in earlier work I have sketched the outlines of a partial ideal theory of gender justice, by specifying three principles of gender justice. It is highly unlikely that these principles will ever be met, and meeting these principles of gender justice may conflict with meeting principles of other types of partial justice, such as justice between parents and non-parents, or what justice for children requires. The only claim that such an account of gender justice makes is that in order for a society to be considered gender just, these principles have to be met.<sup>10</sup>

Saying that ideal theory functions as a mythical Paradise Island may be taken to imply that it has a *direct* guiding function for policy and social change. But so far I have only suggested that ideal theory guides us by telling us where the *endpoint* of the journey lies: it does not necessarily tell us anything about the route to take to get to Paradise Island. In some seas it is dangerous, indeed impossible, to just sail straight in the direction of the destination. For example, in low seas, one needs a precise map of the channels in between the sandbanks—and these channels can make the sailor first head in a very different direction compared to the track that is direct from an aerial view. If sandbanks move over time, an island that was once reachable may no longer be within reach—or at least not until the sandbanks have shifted again. It is not only natural phenomena that may make it impossible for a sailor to sail in a straight line to the island, but also man-made constraints, such as dangerous shipwrecks. Similar dangers hold for attempting to draw straight guidelines for public policies and social action from ideal theories of social justice.

---

<sup>10</sup>Ingrid Robeyns, "When Will Society Be Gender Just?" in Jude Browne (ed.), *The Future of Gender* (Cambridge: Cambridge University Press, 2007), pp. 54-74. Note that it may be the case that gender justice needs to be weighed against other dimensions of justice, or that the price to be paid to reach a fully gender-just society is considered too high; these are questions of nonideal theory, which will be discussed below.

The two other types of normative social justice analysis are both non-ideal analyses. While the boundaries between these two remaining types are fuzzy, for heuristic purposes I will distinguish between nonideal theory on the one hand, and action design and implementation on the other. While the latter builds on theory, it is mainly empirical research.<sup>11</sup>

## 2.2. *Nonideal theory*

In cases in which we are not in a fully just society, we need theory to guide us for two important tasks: first, to be able to make comparisons between different social states and evaluate which one is more just than the other; and, second, to guide our actions in order to move closer towards the ideals of society. The latter is sometimes called the theory of transition. These two main functions of nonideal theory will be discussed in this section.

Before discussing how I understand the two main functions of non-ideal theory, let us first look at how Rawls defines nonideal theory. Rawls holds that nonideal theory

studies the principles that govern how we are to deal with injustice. It comprises such topics as theory of punishment, the doctrine of just war, the justification of various ways of opposing unjust regimes, ranging from civil disobedience and militant resistance to revolution and rebellion. Also included here are questions of compensatory justice and of weighing one form of institutional injustice against another.<sup>12</sup>

To my mind, Rawls is setting us on the wrong foot here, in two ways: on the one hand, he includes issues that may be argued not to belong to the domain of nonideal theories of justice; on the other hand, those questions that would arguably comprise the lion's share of nonideal theory for most contemporary theories of justice are only mentioned in passing, and insufficiently spelled out.

Let us first look at the issues that should arguably not be included in Rawls's list of nonideal theory. To start with, one could question whether just war theory is not simply a different area of moral philosophy, rather than being part of theories of justice. Something similar could be argued for issues of punishment. Suppose we are able to develop and justify an account of the fully just society, and everyone agrees that the principles of justice do indeed give us an account of complete justice. Suppose further that we are able to actually create this society here and now. Then at this point there may be people who, whilst acknowledging that the soci-

---

<sup>11</sup>I use the word "action" rather than "policies," since apart from the government there are also other agents of justice. Justice-enhancing actions do not only include policies, but also activism and public action by civil society groups or by individuals.

<sup>12</sup>Rawls, *A Theory of Justice*, p. 8.

ety is just, violate the ideal principles, for example, by murdering someone or by violating just property rights (for whatever reason they may have, ranging from being blinded by passion, to pure self-interest, or being bored). For these kinds of criminal offenses, we need institutions of criminal justice based on a theory of punishment. I do not think theories of justice need to be terribly concerned with this kind of noncompliance, since it is reasonable to assume that there will always be at least a few citizens who will violate the laws, whether these laws are just or unjust. The notion of full compliance should instead be taken to mean that *under ordinary circumstances*, most people would comply with the principles. Whereas issues of compensatory justice and weighing should be properly dealt with in theories of nonideal justice, issues of war and criminal justice may require separate theorizing and analysis within moral philosophy.

Moreover, most of the more pressing issues of nonideal theory are insufficiently stressed in Rawls's definition of nonideal theory, such as whether the ideal principles of justice need to be adapted when we are theorizing justice in nonideal circumstances, or how to weigh the different principles of justice. Thus, I propose that we put Rawls's definition of nonideal theory aside, and instead focus on what I regard as the two main functions of nonideal theory: first, to enable us to make comparisons between different social states and evaluate which one is more just than the other (this is what Sen has called comparative justice);<sup>13</sup> and second, to guide our actions in order to move closer towards the ideals of society. Although these are analytically two different tasks, in practice they often run together, especially since the latter presupposes the former.

Nonideal theory departs from ideal theory in focusing not on principles of justice in the perfectly just society, but rather on offering us the theoretical foundations for figuring out what we have to do in order to move closer to that society. Nonideal theory should develop the constitutive parts of a theory of justice that are needed in order to bring us one step closer to justice assessments and policy design. Michael Phillips has argued that the principles that are appropriate to the ideal world (in which we assume that everyone would comply with the principles of justice), are not *immediately* applicable to nonideal worlds, such as the one in which we currently live.<sup>14</sup> If pressed on this issue, defenders of ideal theory would not deny this; indeed, they are likely to argue that many, or perhaps most, of the critiques on ideal theory are due to a misunderstand-

---

<sup>13</sup>Sen, "What Do We Want From a Theory of Justice?"

<sup>14</sup>Michael Phillips, "Reflections on the Transitions From Ideal to Non-Ideal Theory," *Noûs* 19 (1985): 551-70. This claim will be further discussed in section 3.3 below.

ing of its limited role.<sup>15</sup> Yet the literature on theories of justice remains remarkably silent on what, then, precisely is needed in order to make the transition from ideal theory to nonideal theory and action design and implementation. Perhaps we need to interpret the ideal-theoretical principles in a context with nonideal circumstances. Perhaps we need to develop a new set of nonideal principles of justice, which are developed by adding layers of relevant facts from the nonideal world to the ideal theory, using the “theoretical resources” that are available in the ideal theory. To the best of my knowledge, there is no systematic and comprehensive account of what a nonideal theory of justice entails, or on which methodological basis it would rest. There are some scattered contributions to the question of how we could go about developing nonideal theory, and also some insightful work within nonideal theorizing about justice, yet this body of literature remains far removed from a systematic account of nonideal theory.<sup>16</sup>

One important part of nonideal theory is the development of principles for comparisons of justice in different social states. These principles would have to tackle the difficult issue of how to weigh different principles and domains of justice, or they may specify priority rules.<sup>17</sup> For example, Nussbaum’s capabilities theory of justice specifies ten domains of capabilities that should be guaranteed to all people by the government as a matter of minimal social justice. She argues that for each individual, the government should guarantee that minimal threshold levels of these capabilities are secured. This is an example of ideal theory: if all individuals have the capabilities at these threshold levels, then, according to Nussbaum’s theory, minimal justice is attained. However, her theory does not answer a number of important nonideal questions. For example, there are many instances in which her theory does not allow us to judge one situation as more unjust than another. Nussbaum fails to tell us how

---

<sup>15</sup>See for example, Adam Swift, “The Value of Philosophy in Nonideal Circumstances,” *Social Theory and Practice*, this issue, pp. 363-87.

<sup>16</sup>For partial answers to the question how we could develop nonideal theory, see Phillips, “Reflections on the Transitions From Ideal to Non-Ideal Theory”; Robert Goodin, “Political Ideals and Political Practice,” *British Journal of Political Science* 25 (1995): 37-56; Harry Brighouse, *Justice* (Cambridge: Polity Press, 2004), chap. 2; and Lisa H. Schwartzman, “Abstraction, Idealization, and Oppression,” *Metaphilosophy* 37 (2006): 565-88. For examples of nonideal theory, some of which is explicitly grounded in ideal theory, see Susan Moller Okin, *Justice, Gender, and the Family* (New York: Basic Books, 1989); Roland Pierik, “Reparations for Luck-Egalitarians,” *Journal of Social Philosophy* 37 (2006): 423-40; Tommie Shelby, “Justice, Deviance, and the Dark Ghetto,” *Philosophy and Public Affairs* 35 (2007): 126-60; Jonathan Wolff and Avner De-Shalit, *Disadvantage* (Oxford: Oxford University Press, 2007).

<sup>17</sup>See Goodin, “Political Ideals and Political Practice,” for the consequences that trade-offs have for the relevance of ideal theory in nonideal circumstances.

the different capabilities should be weighed against each other, or which one we should prioritize in case we can only reach threshold levels for one of them. Suppose we are in situation A, and through policy intervention and social action we can move to only two other social states, B1 and B2. In B1, all capabilities that the individuals had in A are secured at the same level, except they all have higher levels of the capability health, and for half of the population this implies that they are now reaching the threshold level for this capability. In B2, all capabilities that the individuals had in A are secured at the same level, except they all have higher levels of the capability knowledge, and for half of the population this implies that they are now reaching the threshold for this capability. All other things equal, which social state is to be preferred from the point of view of social justice? Nussbaum's theory fails to give an answer to this question, or to sketch out a procedure for deciding on such matters in a democratic fashion. Thus, her theory lacks important parts of nonideal theorizing.

Another important area of nonideal theory is how to make choices between different domains of partial justice theory. For example, theoretical principles of gender justice establish when society will be gender just—but realizing gender justice may be in tension with realizing other demands of justice, for example, what justice requires for children, or what justice between parents and non-parents entails. At present, ideal theory is not performing that task. That may be because in practice all existing ideal theories are to more or lesser degrees partial theories, and thus their incompleteness means that the theory doesn't tell us what to do if different dimensions of justice come into conflict. Depending on how an ideal theory is developed, it may also be the case that it does not have the resources to tell us what to do if different dimensions come into conflict. In either case we need to rely on nonideal theory of justice to spell out how we will make trade-offs between different domains of partial justice. Nonideal theory also entails an analysis of how to make trade-offs between the ideals of social justice and other values, such as efficiency, stability, or sustainability—since these choices inevitably have to be made when considering justice-enhancing policies.

### ***2.3. Action design and implementation***

Ideal and nonideal theories of justice tell us what ideals we are striving for, how different principles of justice should be weighed against each other, how justice needs to be balanced against other values, and how to deal with instances of widespread non-compliance. Yet this is not sufficient for the design of action (including policies); therefore we need the help of social scientists. When designing actions (especially policies), we

also need to take into account a whole range of feasibility constraints and unintended consequences.

Regarding feasibility constraints, it is important to distinguish between those that we have good reason to take as virtually unalterable by society, versus those that are more contingent. These feasibility constraints can be situated on a continuum, from the completely unalterable (such as human mortality or the dependency of human life on the presence of oxygen) to the much more adaptable. For some, the position on this continuum is not fixed over time; some constraints will be unalterable today, yet may become feasible in the future thanks to technological change or a change of societal values and our self-understanding. Surely a century ago people could not have imagined that one day there would be a safe and accessible method of birth control that did not require complete sexual abstinence. Yet this social change arguably has important consequences for a range of moral questions. Similarly, we may not know which constraints that currently seem unalterable will become alterable in the future. For example, the constraint that men cannot become pregnant seems at present rather unalterable, but we cannot preclude the possibility that at some point in the future this would change. Examples of constraints that are clearly much more socially constructed and amendable are that people's social rights depend to a significant extent on the passport(s) they hold, or that societies have a dominant set of social norms that may hamper the realization of principles of justice. When designing policies or justice-enhancing actions, one needs to be aware how amendable certain constraints are.

Unintended consequences are very important in policy and strategy design, and explain why so many well-intended policies do not contribute to the realization of the intended ideals. For example, not taking into account the identity-related sensitivities of the population may produce unintended consequences.

Once we are confident that we have developed and justified the best ideals of justice, that we have successfully complemented and further developed them into nonideal theory, and that we have taken account of feasibility constraints and unintended consequences in action design, we still need one more layer of work before justice can be realized: the stage of the implementation of the justice-enhancing action. At this stage we need to answer questions such as the following: how can we communicate and implement the policies or strategies so as to earn the support of the relevant agents? What aspects of the process of implementation are important in their own right? And, what kind of processes are respectful and democratic, or make optimal use of any untapped knowledge? There exists some kind of "administrative" or "thin" approach to policy implementation that does not take into account the support of the relevant

agents, but virtually no policies can succeed if the relevant agents do not comply with the policies and strategies. This may require processes of participation, information, and involvement of the relevant population, so that the policies and strategies are felt as “jointly owned” by both the policy makers and the affected citizens. If the ideals of social justice require that certain unjust habits or social norms change, then one may have to carefully consider how to set up this critique: will an internal or external critique be most effective, what are the relevant emotional or social-psychological mechanisms at work, and so forth. For example, social justice advocacy stemming from outside the country in which the change should take place, especially if the advocacy comes from countries that are (perceived to be) much more hegemonic, or that have a history of colonial domination, may be perceived as illegitimate by the affected population. This may produce unintended consequences, which may make a justice-enhancing strategy that looks fine on paper entirely ineffective, or even counterproductive. While from a purely intellectual or philosophical point of view the social justice advocacy may be entirely valid and convincing, it nevertheless may have very negative effects if one does not consider the legitimacy and authority of the advocates and implementers. Similar arguments have been made regarding the often countereffective critiques by Western feminist philosophers on gender inequalities in the global South.<sup>18</sup> These kinds of implementation issues raise very difficult questions related to political and identity-related sensitivities and require anthropological insights that go far beyond the professional expertise of most political philosophers. Yet we nevertheless should be aware that social justice-enhancing policy and strategy implementations that do not respect these sensitivities and behavioral responses of those groups are likely to fail, and may even create a social state that is more unjust.

The implementation stage raises still further issues. Sometimes, in order to move towards the ideals of justice, it may be necessary to implement policies and strategies that can be considered “illegitimate,” since in the short term they are creating more injustice in dimension A, while any potential justice gain in dimension B will only be in the medium or long term and is not guaranteed. In other words, one needs to sacrifice justice in one dimension here and now, in the *hope* of gaining more justice in another dimension (and restoring justice in the first dimension) in the long term. Or, a variation on this problem: in the short term we are sacrificing justice in dimension A for group X, *hoping* that

---

<sup>18</sup>Alison Jaggar, “‘Saving Amina’: Global Justice for Women and Intercultural Dialogue,” in Andreas Follesdal and Thomas Pogge (eds.), *Real World Justice* (Dordrecht: Springer, 2005), pp. 37-63.

in the long term this will create a more just situation in the same dimension A for group Y (with the justice shortfall for Y being much larger than for X).

In conclusion, the road from ideal principles to effective justice-enhancing action is long and potentially thorny, and much work is needed before ideal principles can effectively contribute to solving problems of injustice. For ideal theorists of justice, the main lesson to draw is that their work is only one part in a large chain before any change of justice may be reached. If ideal theorists want to contribute to justice-enhancing change in practice, then the point to take home is that although ideal theorizing may be an essential piece of the work to be done, it nevertheless remains only a small fraction of all the work that needs to be done. This prompts questions about its academic status and priority, and the time and other resources we should allocate to this kind of work—questions that I will address in the concluding section of this article.

### 3. Idealizing Assumptions in Ideal Theory

In the previous section, I defined ideal theory as the theory that works out and justifies the principles of justice in a fully just society. Yet in the literature, ideal theory is often confused with theory based on *idealizing assumptions*, in particular, idealizing assumptions that are allegedly bad idealizations. This confusion has nourished many critiques of ideal theory. In order to create some analytical clarity, several questions need to be asked regarding the role of idealizations in ideal theory. What are idealizations? How do idealizing assumptions relate to ideal theory? Are all idealizations in ideal theory unwanted? If not, on what grounds could we distinguish between good and bad idealizing assumptions? And, what are the problems created by these idealizations in the practice of ideal theory?

#### 3.1. What are idealizations?

Idealizations are assumptions that describe certain aspects of a theory differently from how they are in reality.<sup>19</sup> These aspects are often related

---

<sup>19</sup>The account that follows borrows from and builds on discussions at the ECPR joint session, “Social Justice: Ideal Theory, Non-Ideal Circumstances,” Helsinki, May 2007 (see especially Valentini, “On the Apparent Paradox of Ideal Theory,” and the papers by Adam Swift and Zofia Stemplowska in this issue), and a range of definitions that have been presented in the literature, including Onora O’Neill, “Abstraction, Idealization and Ideology in Ethics,” in J.D.G. Evans (ed.), *Moral Philosophy and Contemporary Problems* (Cambridge: Cambridge University Press, 1987), pp. 55-69; Brighouse, *Justice*, p. 19; Schwartzman, “Abstraction, Idealization, and Oppression.”

to human beings, such as their character, motivations, and capacities. Examples are Rawls's assumptions that citizens are free and equal moral persons who have two moral powers (the capacity for a sense of right and justice, and the capacity to form and pursue a conception of the good), and that these citizens are capable of taking part in social cooperation for mutual advantage and desire to do so.<sup>20</sup> Other idealizations concern the nature of society. For example, Ronald Dworkin assumes in his account of justice that society is closed without immigration flows.<sup>21</sup>

Idealizations are simplifications. Often they depict reality as much better than it is.<sup>22</sup> For example, in working out his ideal principles of justice, Dworkin assumes that members of the society have authentic preferences and that their actions and choices are never influenced by prejudice. That assumption makes people into morally better persons than they really are, since we know that in reality our actions are influenced by stereotypes and prejudices.<sup>23</sup>

### 3.2. *How do idealizations relate to ideal theory?*

Most—perhaps even *all*—ideal theories of justice make use of idealizations. In part this is because idealizations are forms of abstractions, and the very nature of theory construction requires us to use abstractions. Sometimes the use of idealizations is necessary to keep the complexity of the theory within manageable boundaries. By introducing idealizations, we reduce the number of parameters that the theory has to deal with. The problem is similar to the solving of a set of equations in mathematics: if there are too many unknown variables relative to the number of equations, then the set of equations cannot be solved and there is no solution to the problem that the set of equations describes. Something similar takes place in the construction of ideal theories of justice; we reduce the complexity by making some aspects of society and of persons simpler, and thereby often better than in reality. In that way we can focus on the essence and get a grip on the complex set of questions.

Another important role for idealizations lies in the goal of ideal theory

---

<sup>20</sup>John Rawls, "Social Unity and Primary Goods," reprinted in John Rawls, *Collected Papers*, ed. Samuel Freeman (Cambridge, Mass.: Harvard University Press, 1999), pp. 359-87, at p. 365.

<sup>21</sup>Ronald Dworkin, *Sovereign Virtue: The Theory and Practice of Equality* (Cambridge, Mass.: Harvard University Press, 2000).

<sup>22</sup>Idealizations need not always present reality better than it is, as we will see below in section 3.4, where I present the example of idealizations that are assuming away the caring dimensions of personhood.

<sup>23</sup>See the vast literature in social and cognitive psychology that has documented these phenomena. For an overview study related to gender stereotypes and prejudice, see Virginia Valian, *Why So Slow? The Advancement of Women* (Cambridge, Mass.: MIT Press, 1997).

itself, namely, to model desirable properties of the ideally just society. Recall that ideal theory aims to give us the principles of justice that would have to be met in a fully just society. In such a society, we wouldn't want people to act on prejudices and stereotypes. It therefore makes sense to assume that in the perfectly just society such prejudice and stereotypes would not exist: people would have authentic preferences and would not be motivated by stereotypes. This is precisely what Dworkin does in his theory of equality of resources.<sup>24</sup> While working out the just distribution of resources, Dworkin assumes that the principles of authenticity and independence would be met. The principle of authenticity ensures that members of the immigrants' society have authentic preferences, while the principle of independence ensures that they do not engage in actions or choices that are influenced by prejudice. Clearly these idealizing assumptions have consequences for the ideal theoretical principles that Dworkin develops. These two principles imply that Dworkin's egalitarian theory is developed in a context in which all socially generated inequalities are assumed away. In his ideal theory, Dworkin uses a *tabula rasa* approach whereby the newly built society has no history of subordination of women, homophobia, slavery, class domination, inequalities in inherited wealth, and other historical processes that have made the playing field unlevel in the real world here and now.<sup>25</sup> Ideal theorists of justice would rightly respond that it is necessary to temporarily bracket details if one is dealing with complex questions of justice in order to keep a grasp on the complexity of the theory. Moreover, these idealizations are arguably theoretically justified at the level of ideal theory, since in a fully just society we would not see any of the inequalities that Dworkin has assumed away when introducing his new society in which people have no distorted preferences.

It is important to understand *why* we idealize. We assume away certain injustices and their causes since at the ideal level these injustices should simply not occur. Examples are prejudices or slavery. If we have strong first principles justifying why these factors can *never* be justified, then we don't need to theorize about the question whether they are just or unjust; in other words, they don't need to be part of the theory. If we all agree that under any conceivable account of justice prejudices create injustices, then we know that at the level of ideal theory we can assume these injustices away by introducing idealizations. In contrast, we cannot introduce the idealization that everyone will earn the same hourly wage, since it is not at all clear that this is what each and every plausible ac-

---

<sup>24</sup>Dworkin, *Sovereign Virtue*.

<sup>25</sup>Roland Pierik and Ingrid Robeyns, "Resources Versus Capabilities: Social Endowments in Egalitarian Theory," *Political Studies* 55 (2007): 133-52, pp. 141-42.

count of justice requires: that is precisely what we need to work out when theorizing about justice.

If all these arguments are sound, then we should conclude that idealizations have a useful, probably even necessary, role to play in the construction of ideal theories of justice. Why, then, is there so much resistance against idealizations in ideal theory? Apart from possible unsound reactions based on an allergy towards counterfactual thinking and highly abstract reasoning, there are also at least two valid reasons. The first is that the transition from ideal to nonideal theory is everything but straightforward, and the more the ideal theory has been built upon idealizations, the farther away it will be removed from offering us clear guidance for the nonideal world. This problem is little discussed in the practice of ideal theory, which may, at least in the eyes of the critics of ideal theory, create the impression that ideal theory is directly relevant to nonideal theory and to action design and implementation. The second problem is that not all idealizations serve a legitimate purpose such as those discussed in this section; there are bad idealizations that cannot be theoretically justified. Since there has been very little discussion in the literature on what distinguishes useful from bad idealizations in ideal theory, the presence of bad idealizations has been extrapolated into a general critique on *all* idealizations in ideal theory, whether they are useful or not. In what follows, we will discuss both problems in turn.

### *3.3. The transition from ideal to nonideal theory*

By its very nature, ideal theory is an enterprise that is different from nonideal theory and justice-enhancing action design. Since ideal theory relies on assumptions that are not met in reality—idealizations—its resulting principles of justice cannot serve as principles for the nonideal world. They need to be adapted or reinterpreted or further developed for the nonideal world. There are several reasons why this is the case, some of which are related to institutional or feasibility constraints.<sup>26</sup> Here I want to concentrate on a different problem, namely, what kind of principles ideal theoretical principles are if they are the results of theorizing that

---

<sup>26</sup>See, for example, Goodin, “Political Ideals and Political Practice,” for some arguments related to feasibility constraints. A rather striking argument illustrating the limits to implementing ideal-theoretical principles in the real world has been developed by Walter Bossert and Marc Fleurbaey, who have mathematically proven that two very plausible luck-egalitarian principles (similar to Dworkin’s principles of endowment-insensitivity and ambition-sensitivity) cannot together be met in a society that meets some minimal and realistic assumptions. To the best of my knowledge, these results have had no impact at all on ideal theorizing about justice. Walter Bossert and Marc Fleurbaey, “Redistribution and Compensation,” *Social Choice and Welfare* 13 (1996): 343-55.

has assumed away some important dimensions of injustice, that is, if the theory includes some idealizations.

Take, for example, Dworkin's principles of endowment insensitivity and choice sensitivity.<sup>27</sup> We cannot implement these principles *directly* in the present world, since in the present world the choices that we make are not as "pure" as they are in Dworkin's theory. Recall that Dworkin assumes that in the ideal just society our preferences are independent of influences by other people, and that we do not make choices and decisions based on prejudice or stereotypes. If we were to implement Dworkin's ideal theoretical principles *directly*, we would not be able to take into account the fact that our choices are so influenced by injustice-creating factors, such as adaptive preference-formation mechanisms. The Dworkinian ideal-theoretical principles of endowment insensitivity and choice sensitivity are constructed based on the assumption that there are no stereotypes and morally relevant preference-formation mechanisms; if we were to implement these without further adaptation for the nonideal world, then we would not be able to account for the fact that our choices are influenced by injustice-generating factors.<sup>28</sup>

Of course, Dworkin has not just developed ideal theory, but has also done work at what could be taken to be the nonideal level. This part of Dworkin's theory is called "the theory of improvement," and would thus do what we are expecting nonideal theory to do. Dworkin's theory of improvement takes the ideal egalitarian distribution of resources and its liberty/constraints baseline as a benchmark and develops a number of proposals for egalitarian improvement in our unjust, nonideal world.<sup>29</sup> The theory of improvement advocates measures to reduce a person's "equity deficit," which is the shortfall between what a person would be entitled to under the ideal egalitarian distribution and her actual situation. An equity deficit consists of two components: a "resource deficit," which is the difference between the quantity of resources that equality of resources would allocate to a person and the quantity that she actually has, and a "liberty deficit," which is the total of liberties guaranteed by the liberty/constraints baseline that are not secured in reality.<sup>30</sup>

---

<sup>27</sup>The following paragraphs are based on joint work with Roland Pierik; see Pierik and Robeyns, "Resources Versus Capabilities."

<sup>28</sup>One could also wonder whether, if such ideal-theoretical principles were implemented in the real world, they would work in favor of the most powerful members of society, those that are not subjected to prejudiced stereotypes or harming preference-formation mechanisms. See the discussion (which concerns ideal theory in general and not Dworkin in particular) in Schwartzman, "Abstraction, Idealization, and Oppression," p. 571.

<sup>29</sup>The liberty/constraints baseline is a set of five principles that Dworkin presents as an integral part of his egalitarian theory, and includes the principles of independence and authenticity. Dworkin, *Sovereign Virtue*, pp. 147-66.

<sup>30</sup>*Ibid.*, p. 164.

Dworkin's theory of improvement does indeed contain detailed elaborations of ways to deal with resource deficits in actual societies, for example, for health care insurance and welfare.<sup>31</sup> But he does not tell us how liberty deficits can be rectified or compensated for. Dworkin argues that the principle of independence seems to be "an appropriate means" to deal with prejudice and thus is a "general feature" of the conception of equality of resources.<sup>32</sup> Yet, he does not develop this principle into actual public policies. Dworkin does not engage with actual and well-known examples of equity deficits, such as those resulting from the gendered structures of society, or from racial prejudice. In chapters 11 and 12 of *Sovereign Virtue* he discusses whether affirmative action is an effective policy against prejudice, and whether it is deemed constitutional by the current members of the U.S. Supreme Court, but he does not discuss these issues in the light of the principle of independence or other aspects of his ideal theory. Dworkin's theory of improvement does not directly guide us in solving issues of sociocultural inequality or cases in which cultural and economic inequalities interact. Before we can solve these questions, we need to make the transition from ideal theory to nonideal theory. One way to do this is to further develop Dworkin's notion of endowment to include social endowments, which would pick up the effects of stereotyping and group-based preference-formation mechanisms that are bracketed by idealizations.<sup>33</sup>

More generally, the problem is not that idealizations are not acceptable at the ideal level, but rather that we need to know how to deal with the idealization when moving to the nonideal level. At the ideal level, the idealizations assume away those alterable causes of injustice that would be entirely eradicated in the fully just society, and we concentrate on deriving and justifying principles regarding the other injustices. But what does this imply for the nonideal world? Do we have to wait on solving the latter until we have reached the former, for example, until we have guaranteed all basic liberties and no one holds prejudiced views?

What would be the appropriate way to deal with ideal theoretical principles at the nonideal level? One possibility would be to first work towards a society in which the injustices that are assumed away by means of idealizing assumptions would no longer exist, before we work on realizing the other principles of justice. But that would be an unfortunate strategy, since we know from research in social and cognitive psychology that many of the causal mechanisms underlying these injustices are

---

<sup>31</sup>Ibid., pp. 307-50.

<sup>32</sup>Ibid., p. 162.

<sup>33</sup>Pierik and Robeyns, "Resources Versus Capabilities," p. 143; Pierik, "Reparations for Luck-Egalitarians."

very persistent. So we can't just wait until the idealizations have materialized in the real world. But we also can't implement ideal theoretical principles in nonideal circumstances, since these are designed for situations under certain conditions that are not met and may therefore have unwanted unintended consequences (such as worsening rather than enhancing justice) when implemented in a world in which these idealizing conditions are not met. In sum, the transition from ideal principles to nonideal circumstances is anything but straightforward.

### *3.4. Bad idealizations*

The second problem that idealizations in ideal theory may cause is that some of them may turn out to be bad idealizations that do not serve legitimate purposes such as those previously discussed. Bad idealizations are those that do not serve to model a certain absence of injustice at the level of ideal theory in cases in which this is theoretically justified, as in the case of Dworkin's assumptions of authentic preferences and the absence of prejudice and stereotyping. Bad idealizations amount to ignoring the existence of certain forms of injustice that need to be theorized rather than simply ignored in the theorizing. Often this amounts to leaving out aspects of life that are more relevant for some groups in society than for others. These idealizations are often not explicitly indicated in the form of assumptions or background conditions, but rather creep into the theory and require careful reading in order to become visible.

An example is the idealization of the conception of the person by assuming that he is not dependent for care upon others, nor constrained in his actions and plans of life by caring duties. Idealizations related to the need to receive and give care are not the kind of idealizations that are justified as appropriate on the grounds that in the fully just society these needs would no longer exist, as is the case of prejudice and stereotypes. Rather, the inevitable fact of human dependency on the care of others makes the just distribution of care (both at the receiving and giving end) an important aspect that a theory of justice, both at the ideal and the non-ideal level, should work out. The fact of human dependence on the care of others is like the fact of human mortality: both are facts that no serious comprehensive theory can ignore. Care is not something that can be assumed away through idealizations; rather, it has to be confronted upfront, so that the theory can help us to work out what in a just society the principles would be that would regulate just caring. Introducing a conception of personhood that assumes away care is a bad idealization, since it cannot be justified as an idealization that reflects an ideal principle of justice (as in the case of idealizations regarding the absence of stereotypes).

One could object to this argument as follows. Idealizing away care

would be an example of bad ideal theory only if the theory aspired to say something about care, but not if the theory is a partial theory intended to work out the principles of some other domain of justice. Yet this can only be a viable defense if there are no significant spillover effects from the domains in which care operates to the domains for which the partial theory is intended to work out ideal theoretical principles. For example, Rawls's theory of justice aims to work out the principles of political justice for the basic structure of society; but his neglect of care could only be justified if he were to first show that care is morally irrelevant to issues of political justice and the basic structure of society. The arguments presented in the literature on justice and care make it rather doubtful that such an argument could be successfully developed.<sup>34</sup>

These bad idealizations can perhaps also explain why some critics of ideal theory argue that it serves an ideological function.<sup>35</sup> Take again the example of care. In present-day societies, care is not equally distributed between all individuals: women perform much more care than men (both paid and unpaid), and among the paid careworkers, immigrants, lower-class women, and women of color are overrepresented. Therefore, if a theory of justice introduces an idealization that brackets away care, for example by introducing a conception of the person which assumes that people are fully independent (i.e., have no dependents and are not dependent for their well-being on care delivered by others), or by excluding from the scope of justice those institutions in which the distribution of care is to a large extent determined (e.g., the family), then these idealizations will be biased against the groups who are arguably treated unjustly with respect to the distribution of care in our nonideal circumstances. This idealization works to the advantage of those who are benefiting from the current unjust arrangements related to the distribution of care, and in that sense it can be argued that it creates an ideological bias in the theory, even if there is not the slightest intention of introducing such a bias on the part of the theorist.

### 3.5. Implications

If the above analysis is sound, then at least three things follow. First, ideal theorists should *much more explicitly* recognize the limitations of ideal theory, and warn their readers that nothing automatically follows from their theory for the real world. In practice, ideal theorists generally

---

<sup>34</sup>See, e.g., Eva Feder Kittay, *Love's Labor: Essays on Women, Equality, and Dependency* (New York: Routledge, 1999); Okin, *Justice, Gender, and the Family*; Nussbaum, *Frontiers of Justice*; Diemut Bubeck, *Care, Gender, and Justice* (Oxford, Clarendon Press, 1995).

<sup>35</sup>For example, Mills, "'Ideal Theory' as Ideology."

don't do this. Instead they sometimes play with real-life examples that create the false impression that their theories could be applied in non-ideal circumstances, where the idealizations are not a reality. I think that the dangers of not being much more explicit about the kind of theory one is producing, and whether or not it can be applied to the real world, are dramatically underestimated. These dangers include potentially harmful policies and actions in the real world, but also a counterproductive development in the literature on theories of justice, where scholars are talking at cross-purposes and where limited time and energies are directed towards developing critiques that are missing the mark.

In the absence of such explicit discussion of the limits of ideal theorizing, it is not surprising that several critics of ideal theory have criticized ideal theory for its unacceptable implications in the real world. The most influential example of such a critique of ideal theories is probably Elizabeth Anderson's critique of luck egalitarianism.<sup>36</sup> Defenders of ideal theory are right, in my view, when pointing out that Anderson is not criticizing ideal theory for the kind of theory it is, and that therefore her critique is misdirected.<sup>37</sup> Yet, I would also argue that in their practice ideal theorists should give warning much more explicitly about the limitations of their theory, and what (if any) the implications are for the real world. The misdirected nature of these critiques is in that sense a shared responsibility.

The second implication concerns the importance of thinking carefully about whether the idealizations that are introduced can be justified at the ideal level. Idealizations that serve to model the absence of injustices in the fully just society are acceptable at the level of ideal theory; idealizations that do not meet this function, and instead introduce ideological biases such as those that are implicit in the concept of the person, are bad idealizations and should not be allowed in the construction of ideal theories.

Third, this section has once again highlighted the importance of non-ideal theorizing of justice. Ideal theory surely has an important role to play, but is far from sufficient, since there is no straight bridge between ideal theory and guidance of actions. We need nonideal theory, and we have to examine whether we can use the resources offered by ideal theory to work out a set of principles for how to proceed in situations in which idealizing assumptions are not met. Sometimes nonideal theory is taken to be a straightforward application of ideal theory, yet the above analysis has also highlighted that nonideal theory has a much more important role to fulfill in normative theorizing of justice.

---

<sup>36</sup>Elizabeth Anderson, "What is the Point of Equality?" *Ethics* 109 (1999): 287-337.

<sup>37</sup>See, e.g., Swift, "The Value of Philosophy in Nonideal Circumstances."

#### 4. Conclusion: Ideal Theory in Practice

In this article I have analyzed the relations between ideal theory and idealizations, and I have discussed what I regard as the pitfalls of certain forms of ideal theory. The conclusion I have reached is that ideal theory plays a limited role: it looks like the Paradise Island where we ideally would like to be, but it does not tell us how to get closer to the island. That work has to be done by nonideal theory, and justice-enhancing action design and implementation.

Perhaps there is a tacit understanding among ideal theorists that everyone knows these limitations. Indeed, some would even say that one has to be pretty uninformed not to know these limitations. Yet if that is the case, why do we see the publication of a significant number of articles that do not acknowledge these limitations and pitfalls while using a writing style suggesting that the theory is of direct guiding use for justice-enhancing change? Moreover, as I highlighted in section 2, normative work on social justice encompasses a range of disciplines and fields. If ideal theorists want to produce theories that are action-guiding in the real world, and want to avoid the dismissive reactions of nonideal theorists or scholars working on effective justice-enhancing strategies that these ideal theories are of no use in reality, then they have to be much more upfront about the limitations of ideal theory, and invest much more time and effort into working out how ideal theory can be developed into nonideal theory and ultimately into action design and implementation.

Much of contemporary post-Rawlsian theoretical work on justice has proceeded in this fashion: first, by assuming that ideal theory is justified since it would be fundamental for nonideal theory; and second, by limiting their work to ideal theory, often without telling the readers where to look for the nonideal work that is needed to provide a solution to the real-world problems, and without warnings about the strict limitations of what the ideal theoretical analyses imply for the real world. In addition, while ideal theorists may be quick in pointing out that policy proposals have no sound philosophical principles, and that therefore applied scholars should pay more attention to their theoretical work, they often do not seem to think that symmetrical scholarly obligations rest on them to ask how useful their ideal theories are for real life problems or what kind of theory is needed in order for it to significantly contribute to enhancing justice. Ideal theorists would do well to acknowledge, in their scholarly practice, these limitations of ideal theory—at any rate more than is currently the case.

What do we need in order to take this debate forward in a constructive manner? First, we need to sort out the proper definitions and descriptions of ideal and nonideal theory. This paper has made a modest contribution

to this need to sort out their precise goals, by providing a simple typology of normative work on social justice. Second, part of the confusion in the literature is based on the fact that ideal theory is often confused with theory based on idealizations. On the one hand the above analysis shows that these are two different issues, yet on the other hand in practice they often run together. We need to further investigate their relationship. Third, we need to further analyze and debate in what sense the role of ideal theory is limited. Fourth, we need to work through more cases of flawed idealizations, in order to better understand when and why idealizations are bad or unwanted, and when and why they are useful. We also need to understand better under which conditions principles of ideal theory are countereffective if one tries to implement them in the real world. Fifth, *much* more work needs to be done on developing nonideal theory, and on better understanding what is needed for nonideal theory to provide an effective bridge between ideal theory and action design and implementation. Sixth, we need to look critically at how we are socializing students: do we stress the importance of the aesthetic appeal of theories, or rather stress their direct or indirect practical use? Finally, we may need to question the current hierarchical status of ideal theory. We have to ask whether the professional culture within political philosophy and the incentive structures in academic institutions, including the reputations of journals and implicit and explicit value judgments about different kinds of theory (i.e., more versus less abstract “pure,” formal, and so forth) steer us disproportionately into the direction of ideal theory. I contend that this is the case, but would be more than happy to be proven wrong.<sup>38</sup>

**Ingrid Robeyns**

Department of Political Science  
Radboud University Nijmegen  
i.robeyns@fm.ru.nl

---

<sup>38</sup>Earlier versions of this paper were presented at the ECPR joint session on “Social Justice: Ideal Theory, Non-Ideal Circumstances,” Helsinki, May 2007, at the Nijmegen Political Theory Workshop, September 2007, and at the Dutch Research School for Practical Philosophy (OZSE), October 2007. For comments and discussions I would like to thank Harry Brighouse, Nick Ferreira, Pablo Gilabert, Bob Goodin, Hugh Lazenby, Roland Pierik, Amartya Sen, Zofia Stemplowska, Adam Swift, Laura Valentini, Wibren van der Burg, Marcel Wissenburg, the editors of this journal, and an anonymous referee. Financial support of the Netherlands Organisation for Scientific Research (NWO) is much appreciated.