1. Introduction

In the post-Rawlsian literature on theories of justice, ideal theory makes up the lion share of work being done by mainstream political theorists and philosophers. John Rawls clearly understood his work as ideal theory, which he described as a conception of a fully just society. Rawls acknowledged that, with the exception of his analysis of civil disobedience, he would not pursue non-ideal theory. But clearly Rawls thought he prioritised 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 say that “the nature and aims of a perfectly just society is the fundamental part of the theory of justice.” (Rawls 1971/1999: 8). Rawls believes that non-ideal theory cannot be done without first working out ideal theory. Post-Rawlsian theories of justice have broadly endorsed this view on the relation between ideal and non-ideal theory, and the philosophical literature on social justice has been mainly been concerned with working out, revising, testing, and refining ideal principles of justice.

In the last years, however, a growing number of political philosophers have expressed their worries about the nature of ideal theory and its development and use in theories of justice. Jo Wolff (1998:113) has argued that “ideal thinkers who want to have some..."
impact on reality should pay more attention to issues of transition.” Amartya Sen (2006) considers ideal theory (which he calls ‘transcendental theory’) neither necessary nor sufficient to guide justice-enhancing policies. He strongly criticised 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” (Sen 2006, 235, 238). Charles Mills contends that ideal theory “is really an ideology, a distortional complex of ideas, values, norms, and beliefs that reflect the nonrepresentative interests 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” (Mill 2005: 172, italics in original). Colin Farrelly (2006: 2) has also staged a fierce attack on ideal theory, claiming it to be “inherently flawed.”

What is one to make of these strong criticisms on ideal theory? It is striking that in the literature on theories of social justice terms such as ‘ideal’, ‘ideal theory’ or ‘non-ideal theory’ are used in widely and conflicting different ways. The first task of this paper will therefore be provide an attempt at clearing 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 (section 2). I will then distinguish ideal theory from theory relying on idealising assumptions and discuss their relationship. I will argue that while both ideal and non-ideal theory can be idealizing, in practice ideal theory is more prone to being idealizing than non-ideal theory (section 3). Section 4 then assesses the necessity, usefulness and limitations of ideal theory for non-ideal theory and the solving of problems of justice. I will argue that Amartya Sen is correct when arguing that ideal theory is not sufficient or always necessary, but is mistaken in not allocating any contributing role to ideal theory for non-ideal theory and social justice enhancing changes. Section 5 concludes that the role of ideal theory is rather limited and that this should be more acknowledged by ideal theorists – not just in their theory but also in their practice. The paper wraps us by concluding that ideal theory does have a role to play, but that more attention should be paid to analysing and correcting its strongly idealizing assumptions, and that much more time and energy should be spend on the non-ideal parts of normative social justice analysis.
2. Normative social justice analysis: a typology

To start, let us spell out the different types of normative social justice research that we may distinguish, from the most ideal level to actual implementation. There are, in my view, 3 different kinds of research that need to be distinguished:

(a) Ideal theory. The aim of ideal theory is to work out the principles of justice that should govern an ideal 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” (Rawls 1971/1999: 8). When defending and justifying the ideal principles of justice, we assume full compliance with those principles. However, the often heard shortcut among political philosophers that “ideal theory=full compliance” is not very accurate and potentially misleading, since full compliance may also hold with principles of justice of which it is known and accepted that they do not lead to a just society. When defining ideal justice, it should therefore be stressed that it is not only about full compliance with any kind of principles of justice, but full compliance with those principles of justice that are 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. Examples of ideal theory that are aspiring to be comprehensive, or approaching comprehensiveness, are Rawls’s work on justice, or Dworkin’s theory of equality of resources. Partial ideal theory can be partial in several ways. First, it may be partial since 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 (2003, 2006) capability 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. 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

---

3 I am putting aside non-normative work on social justice, such as empirical research on what views about social justice are held by ordinary people (e.g. Miller 1992).
telling us when gender justice will be achieved, while remaining silent on all other domains of justice. Or ideal theory may be partial in a geo-political 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, 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 by threshold level justice in health within a society.

What is the goal of ideal theory so defined? Ideal theory functions as a lighthouse: it tells us in which direction 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’ view, be considered fully just. The same holds for partial theory. For example, in an article entitled “When will society be gender just?”, I have sketched the outlines of a partial ideal theory of gender justice, by specifying 3 principles of gender justice (Robeyns 2007). I argue in that paper that we won’t have reached a state of full gender justice in any particular society, community, area, or indeed the world, before these principles are met.

Saying that ideal theory has a lighthouse function 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 the lighthouse. In some seas it is dangerous, indeed impossible, to just sail straight into the direction of a lighthouse. One needs a precise map of the channels in between the sandbanks – and

---

4 While at first sight Rawls theory can be considered nearing a comprehensive theory of justice, his later clarifications clearly indicated the way in which even Justice as Fairness is partial: it is limited to a closed society, and only provides a political conception of justice, bracketing questions of justice within the family and other private association. Moreover, Justice as Fairness has not worked out (or, in Rawls’s terms, ‘postponed’ a number of questions of justice) such as what justice for the disabled requires. I will discuss this briefly below.
these channels can make the sailor first head into a very different direction before getting onto the track which is direct from an aerial view. If sandbanks move over time, a lighthouse that was once reachable may even no longer be within reach – or at least not until the sandbanks have shifted again. As I will argue below, the same danger holds for attempting to draw straight guidelines about public policies and social action from ideal theories of social justice.

The two other types of normative social justice analysis are all non-ideal thinking. While the borders between these two remaining types are fuzzy, for heuristic purposes I will distinguish between non-ideal theory on the one hand (which in itself can be distinguished into different kinds), and action design and implementation.\(^5\)

(b) Non-ideal theory. While remaining at the level of theory, non-ideal theory of justice relaxes the two fundamental characteristics of ideal theory: full compliance, and the ideal nature of the principles (that is, the fact that if realised, those principles will bring us a fully just society, or justice in a particular domain).

Take the issue of full compliance first. Rawls (1971/1999: 8) holds that non-ideal 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.”

I believe it is important to distinguish issues of punishment and war from those of compensatory justice and weighing. Suppose we are able to develop and justify the fully just society, and everyone agrees that the underlying principles of justice do indeed create 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 society 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

\(^5\) I use the word ‘action’ rather than policies since there are other agents of justice than only the government, and thus justice-enhancing actions do not only include policies, but also activism and public action by civil society or by individuals.
criminal offences to the principles of justice, we need institutions of criminal justice based on a theory of punishment. I do not think theories of justice need to worry about this kind of non-compliance, since it is reasonable to assume that there will always be citizens who will violate the (just) laws. The notion of full compliance should instead be taken to mean that under ordinary circumstances, most people will comply with the principles. Whereas issues of compensatory justice and weighing should be properly dealt with in theories of non-ideal justice; issues of war and criminal justice should, in my view, not be classified under non-ideal theories of social justice, but rather require separate treatment.

Thus, I think Rawls is setting us on the wrong foot by his listing of issues of non-ideal justice. Moreover, most of the real pressing issues of non-ideal theories of justice are not dealt with in Rawls’s work. In part this is because the priority rule in his (ideal) principles of justice have brushed away, at the level of ideal theory at least, the need to think about weighing the different social primary goods. Of course, Rawls had a theoretical justification for this priority rule; the priority of the principle of liberties over the second principle was motivated by the fact that the liberties “cannot normally be sacrificed to gain greater social welfare, or for the sake of perfectionist values; and this restriction is, practically speaking, absolute (Rawls 2007: 12). But in practice this means that Rawls has found an easy solution for not having to deal with one of the most difficult issues of theories of justice – the question how to weigh different dimensions of the metric. This make the usefulness of justice as fairness under non-ideal circumstances of partial compliance strictly speaking rather limited, since we are not allowed to mobilise the second principle (equality of opportunity and the difference principle) as long as the first principle is not guaranteed.

The second way in which non-ideal theory departs from ideal theory is by focussing not on spotless 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. Non-ideal theory should also develop these constitutive parts of a theory of justice that are needed in order to bring us one step closer to justice assessments and policy design. This requires principles for comparisons of justice in different social states (Sen 2006). These principles may tell us how to weigh different ideals and domains of justice, or may specify priority rules (as in Rawls’ justice as fairness).
For example, Nussbaum’s theory of justice specifies 10 domains of capabilities, which should be guaranteed to all by the government as a matter of minimal social justice (Nussbaum 2003, 2006). She argues that for each individual the government should make sure that minimal threshold levels of these capabilities are secured. This is clearly 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 non-ideal questions. For example, her theory does not allow us to judge one situation as more unjust than another, since Nussbaum fails to tell us how the different capabilities should be weighed against each other. 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 individuals have the same capabilities as in A, 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 individuals have the same capabilities as in A, 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 does not tell us, since hers is a purely ideal theoretical account, and lacks important non-ideal theorizing, such as how to make trade-offs between different capabilities.

Another important area of non-ideal theory is how to make choices between different domains of partial justice theory. For example, theoretical principles of gender justice tell us 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. In that case a non-ideal theory of justice needs to spell out how we will make trade-offs between different domains of partial justice.

Thirdly, non-ideal theory needs to tell us 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 and other types of action.

(c) Action design and implementation. Ideal and non-ideal theories of justice are still not sufficient for the design of action: for that we need to know much more than
only what ideals we are striving to, how different principles of justice can be weighed against each other, how justice needs to be balanced against other values, and how to deal with instances of non-compliance. 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 where we have good reason to take them as virtually unalterable by society, versus those that are more contingent. Examples of the first are that there are 24 hours in a day, that men cannot be pregnant, or that by genetic misfortune some people are much more prone to fatal diseases than others. Examples of the second 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, which may hamper the acceptance and feasibility of the realisation of principles of justice. One reason for distinguishing between these types of feasibility constraints is that the first type are hard constraints, whereas the latter type may be changed, but that this may take time and require very smart strategies.

Unintended consequences are very important in policy and strategy design, and explain why so many well-intended policies end up creating a nightmare or, at best, did not contribute to 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 complemented and further developed them with non-ideal theory, and that we have taken account of feasibility constraints and unintended consequences in the action design, we still need one more layer of work before any justice can be realised: 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 command support from the relevant agents? What aspects of the process of implementation are important in their own right, for example 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 from 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 that are at work, and so forth. For example, several religions endorse practices that can be considered unjust to women. However, most religions will not accept critiques on these practices from individuals or groups that they consider as hostile to their religion. For example, it does not help if an atheist woman, such as Ayaan Hirsh Ali, who has repeatedly stated that she believes that the whole of Islam is a backward and an irreducibly undemocratic religion, tells Muslims that they should be less patriarchal. Even Muslim women, quite understandable, experience this as such a strong attack on their group, that they develop a defensive attitude and thereby reinforce their Muslim identity over their female identity. If social justice for women of minority religions, or for gays in non-liberal cultures, requires changing social norms and patriarchal habits, one needs to think very carefully about the appropriate implementation strategies. Similar arguments can be made about 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 powerful in international geo-political terms, or have a history of colonialist domination. While from an intellectual or a philosophical point of view the social justice advocacy may be entirely valid and convincing, it may have completely opposite effects if one does not consider the legitimacy and authority of the advocates and implementers. Similar arguments have been made regarding the often counter-effective critiques by western feminist philosophers on gender inequalities in the global South (Jaggar 2005).

These kind of implementation issues raise very difficult questions related to political and identity-related sensitivities and anthropological insights that go far beyond the scope of this paper; nevertheless even theorists of justice should be concerned whether social justice-enhancing policy and strategy implementations that do not respect these sensitivities and behavioural responses of those groups, are likely to fail or even to lead us to a social state that is even more unjust.
The implementation stage raises still other difficult 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 worsening justice in dimension A, while any potential justice gain in dimension B will only be in the medium or long term, and is even not guaranteed. In other words, one needs to sacrifice justice in one dimension in the here and now, in the hope of gaining more justice in another dimensions (and restoring justice in the first dimensions) in the long term. Or, a variant on this problem is that in the short term we are sacrificing justice in dimension A for group X, hoping that 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 is much larger than for A). An example may clarify. Suppose that liberal society has an immigrant group which practices gender-segregation and where women’s opportunities are severely constraint. One of these women-unfriendly constraints is that there is an almost absolute expectation from married women to become mothers of a several children, and to make motherhood into their only occupation in life. In such a society, a (local) government may decide to set up a toddler playgroup for mothers only, and use it as a forum to mix these immigrant women with more ‘liberal’ women, in the hope to provide these immigrant women with information and networks that will help them to gain better opportunities and to ‘empower’ them to stand up against the patriarchal traditions in their community. Such a mothers-only playgroup is deeply troubling for liberal gender-equalitarians, since it deprives fathers of the opportunity to go with their toddlers to these playgroups, and it hampers both fathers and mothers to practice gender-equalitarian parenting. The short-term result is probably a worsening of justice. But this short term sacrifice is made as a cost to be paid for an expected long-term justice-gain that is arguably much more important than the short-term justice-cost. Since it is an expected gain it remains a decision under uncertainty; but nevertheless we can see why such policy implementation decisions, which in the short term are not enhancing justice, may be the right thing to do in the long term, or put differently, may be a sacrifice worth risking.

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 contributively to solving problem of injustice.
3. Idealizing assumptions in ideal theory

The definition of ideal theory proposed in the previous section is how I believe ideal theory should be defined; in contrast, ideal theory is often confused with theory based on idealizing assumptions. These two issues need to be distinguished, if only because they pose different problems and questions.

What are idealizing assumptions? In her discussion on abstraction and idealisation of agents in ethics, Onora O’Neill (1987: 55) points out that idealisation can take two forms: much of what is true of human agents is omitted in some accounts, but in addition much that is false of human agents is added. O’Neill’s definition of idealisation can be generalised not just to apply to human agents, but also to other parts of the premises in theories of justice. Let us call theories which rely in some crucial respect on such idealised assumption idealising theories.

Take as an example Dworkin’s theory of equality of resources. There are at least two strongly idealising assumptions in Dworkin’s theory: the principles of authenticity and independence with are assumed to be met before Dworkin’s egalitarian auction takes place. 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. As Roland Pierik and I have argued, these two principles imply that Dworkin’s egalitarian theory is developed in a context in which all socially generated inequalities are assumed away (Pierik and Robeyns 2007: 141-2). In his thought experiment, 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 unlevelled the playing field in the real world here and now.

Ideal theorists of justice are generally quick in pointing out that it necessary to temporarily bracket details if one is dealing with complex questions of justice, but once we know the ideal principles of justice, and we try to work out the non-ideal principles for our actual societies, the bracketed issues will be taken on board again. This part of Dworkin’s theory is called ‘the theory of improvement’, and would thus do what we are expecting non-ideal 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 non-ideal world. 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 their 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 they actually have, and a ‘liberty deficit’, which is the total of liberties guaranteed by the liberty/constraints baseline that are not secured in reality (Dworkin 2000: 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 (Dworkin 2000: 307-350). 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 (Dworkin 2000: 162). However, 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 of 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 American Supreme Court, but he does not discuss these issues in the light of the principle of independence or other aspects of his ideal theory. Our conclusion is that Dworkin’s theory of improvement does not directly guide us in solving issues of socio-cultural inequality or cases where cultural and economic inequalities interact (Pierik and Robeyns 2007: 143). The problem, moreover, is not just that Dworkin’s theory fails to directly guide us about how to deal with socio-cultural inequalities in the real world, but also that the strong idealizing assumptions in his theory may lead to flawed and misleading guidance regarding those inequalities that are treated within his theory. The debate between Dworkin and Andrew Williams on how Dworkin’s egalitarian theory would judge gender inequalities clearly showed, at least to my

---

6 The liberty/constraints baseline is a set of five principles which Dworkin (2000: 147-166) presents as an integral part of his egalitarian theory, and include the principles of independence and authenticity.
mind, that his theory is unsuited to deal with real life injustices where the socio-cultural and socio-economic dimensions of inequality interact (Williams 2003; Dworkin 2003; Pierik and Robeyns 2007). Dworkin gives us ideal principles of justice, but these are derived from a model in which such heroic assumptions are made about human psychology (cfr. the principles of independence and authenticity), that we can say with confidence that it is very unlikely that we will live in such a world any time soon.

This discussion of Dworkin’s equality of resources illustrates what idealizing assumptions are, and how such idealizing theories may misguide us when thinking about what to do in the here and now. In my view, Dworkin’s theory is not only one of the most ideal theories of justice (due to his tabula rasa approach), but also one of the most idealised theories (due to a number of strongly idealised assumptions on which his theory is based, such as those embodied in the principles of independence and authenticity).

Ideal theories of justice differ in their degree of idealization. Some contain hardly any idealization at all, and are in fact in part based on a critique of the idealising character of other ideal theories. Martha Nussbaum’s capability approach is a good example: while her theory is certainly an ideal theory (depicting an account of what minimal justice would require), there are no immediately obvious strong idealizing assumptions in her work. Instead, one of her starting points is that other theories, such as Rawls’s justice as fairness, bracket human dependencies and vulnerabilities. Rawls work does indeed start from certain idealisations. His early work assumed that the family would fall outside the scope of justice, since family members behaved altruistically towards each other. This is a clear example of an idealising assumption, as all the empirical evidence on marital violence and intra-household inequalities in access to resources and decision making have shown. Another idealisation is Rawls’s assumption of closed societies – for example, global (legal and illegal) migration continues to illustrate, people have always crossed borders, and a theory of justice needs to question their moral and political legitimacy.

The comparison between idealizing assumptions in Rawls and Dworkin points at another issue: the question is not so much whether a theory is based on idealizing assumptions, but rather whether these idealizing assumptions are playing a foundational or constitutive role in the construction of the principles of justice. In
Dworkin’s theory, this is clearly the case, since people’s preferences play a key role in deciding what the just distribution of resources is (since this distribution should be ambition sensitive and reflect people’s willingness to pay and this is revealed through people’s preferences). The idealized assumptions in Dworkin’s theory thus inescapably affect the principles of justice since the very construction and justification of the principles are based on these assumptions. This may be different for Rawls’s principles of justice, since the idealising nature of his theory lies primarily in the fact that he restricts the scope of this theory to a closed society and to people who function within the normal range of abilities (i.e. severely physical and mentally disabled citizens are excluded from his theory). The question that needs to be asked is whether inclusion of these groups into the scope of justice would alter the ideal-theoretical principles of justice. The answer of political philosophers to this question differs: Brighouse (2001) believes that the physically disabled can be integrated into justice as fairness since it only requires the modification of non-core assumptions of justice as fairness, but he is worried about the accommodation of the mentally disabled; Kittay (1999) and Nussbaum (2006), in contrast, think that the disabled, and more generally issues of dependency and care, cannot be accommodated in justice as fairness. If their arguments (which I will not further discuss here) are correct, and if we hold the view that the disabled should not be excluded from the scope of justice, then this would imply that Rawls’s idealizing assumption may proof problematic for his ideal theory. On the other hand, if one can accommodate the disabled within justice as fairness, and extent the principles to the global realm, than Rawls’s idealizing assumptions are not constitutive of his principles of justice, and hence one could remove the idealizations without having to abandon the ideal principles of justice.

Summing up, it may be helpful to distinguish between idealisations where removing the idealisations doesn’t affect the principles of justice, from idealising assumptions which are constitutive for the principles of justice. The former may, perhaps, be justified on grounds of simplifications (or, as Rawls put it, to focus on the key issues first, and refine and extent the theory later by taking the bracketed issues on board). The latter, in contrast, cannot be justified, since they lead to flawed principles of justice. I thus conclude that all idealising assumptions need to be treated with suspicion, but that only those that are constitutive of the principles of justice should never be allowed in the construction and justification of ideal theories of justice.
In the introduction of this paper I mentioned that in the literature on theories of social justice terms such as ‘ideal’, ‘ideal theory’ or ‘non-ideal theory’ are used in widely and conflicting different ways. Some recent criticisms on ideal theories are, in fact, critiques on their idealising character. For example, Mills’ (2005) fierce attack on ideal theory suffers from insufficiently distinguishing between ideal theory proper and the idealizing nature of (some) ideal theory. Similarly, Farrelly (2007) claims that ideal theory is “inherently flawed”, but comes to this conclusion based on critiques of the idealizing nature of two ideal theories, rather than by arguing why ideal theory properly understood would be inherently flawed. Put differently, he doesn’t tell us whether he thinks that ideal theory without idealization is possible (he suggests that all ideal theory is also be idealizing), and if non-idealizing ideal theory were possible, whether it would have any use and legitimacy. Thus, it follows that Farrelly does not succeed in showing that ideal theory is inherently flawed; at best he may succeed in showing that idealized ideal theory is flawed.

If critiques of ideal theory are so often mixed up with critiques of their idealising character, then we may wonder whether they are in practice one and the same thing: perhaps all ideal theory is idealising, and all idealising theory is ideal theory? I don’t think this is the case. First, I think that both ideal and non-ideal theories of justice can be idealizing. Here’s an example of how an idealising non-ideal theory of justice would look like. Suppose we agree that justice requires that Nussbaum’s ten capabilities need to be met at certain well-defined threshold levels. This is our ideal-theoretical account of justice. Non-ideal theory then requires, among other things, that we develop an account of how, under conditions of scarcity, trade-offs would be made between these ten capabilities and/or between the capabilities of different people. Recall that Nussbaum’s list includes a number of items that are regarded by many citizens as sensitive or contested issues, some on which they are unwilling to

7 Moreover, I believe that Farrelly’s critique of Dworkin is flawed. He argues that (1) Dworkin only considers two types of brute luck, (2) he does not take the cost of insuring against misfortunes into account, and (3) that Dworkinian applications treat misfortunes in isolation and fail to recognize that the list of misfortunes is enormous and that we may therefore not be willing to compensate for all of them (like fertility treatment). However, Dworkin’s theory could easily be extended to account for these problems without changing any of his principles or their basic justification. The islanders in Dworkin’s theory would be given X clamshells, and would know that there are Y misfortunes, and that we will have to pay a certain percentage of our clamshells as the cost of insuring. Under these (more realistic) conditions, we may decide not to insure against certain forms of misfortune. But the basic principles of Dworkin’s egalitarianism, namely the principles of endowment-insensitivity and ambition-sensitivity, remain intact.
compromise for religious or ideological reasons (such as reproductive rights). In answering the challenge of trade-offs, we could develop an idealising account of deliberative democracy, where people would have non-distorted preferences, full information, are sufficiently rational and reasonable, and so forth. Since in reality democratic deliberation seldom meets these conditions, this would be an idealised non-ideal development of a theory of justice.

However, while this example illustrates that both ideal and non-ideal theories of justice can be idealising, I think that ideal theories of justice are more likely to be based on idealising assumptions than non-ideal theories of justice.

There are at least two reasons for that. Firstly, ideal theory generally assumes full compliance. This in itself is an idealising assumption, which may make it in fact particularly hard to develop the ideal theory into a non-ideal theory and action proposals. Rawls’s theory allows us to illustrate this point. Rawls’s theory of justice can serve as an illustration of why the assumption of full compliance is closely linked with idealizing assumptions. In a very fine article on the limits of ideal theory, Michael Phillips (1985) argues that ideal theory should be build on the assumption that that most of us are psychologically capable of doing what the assumption of full compliance demands. For example, if principles of ideal theory would demand that married/partnered people should be allowed to have intimate relationships with others, then strict compliance would mean that we would all accept that as a matter of justice we should support the organisation of our social institutions that support multiple intimate relationships. Phillips (1985: 558) rightly argues that such an ideal theory would be problematic since it would ask something which most people are incapable of doing. Thus, he believes that ideal theory should not demand, “What principles ought a society to adopt for the purpose of designing its institutional structure on the condition that all of the institutions in that society conform to these principles?” or “What principles would an ideally structured society publically acknowledge and enforce to govern the behavior of its members, on the assumption that all members comply with those principles?” but rather “Given the level of compliance that can reasonably be expected, what principles ought a society to adopt for the design of social institutions?” and “Given the level of compliance that can reasonably be expected, what principles ought a (purely or impurely) ideally structured society publically acknowledge and enforce to govern the behavior off its
members?” (Phillips 1985: 553-4). In short, strict compliance itself may be idealizing to a degree that the ideal principles may not be applied to the non-ideal circumstances where this strict compliance doesn’t hold.

The second reason why ideal theories are more likely to be based on idealising assumptions than non-ideal theory is that there is no need to have a dialogue with empirical research, as is the case with non-ideal theories. There may well be epistemological effects, since engaging with empirical research (even if it does not constitutively affect the theorizing) is likely to prevent theorists from producing theories for the sake of the theories themselves – rather the *ultimate* (though therefore not necessarily direct or immediate) problem-solving relevance for the non-ideal and messy world will function as an epistemological constraint on the work of the theorist.

Concluding this section, I think there are three important issues. First, it is important to distinguish between ideal theory and theory based on idealizing assumptions. Second, idealizing assumptions that are constitutive in the construction and justification of the principles of justice will lead to flawed principles. Third, while both ideal theory and non-ideal theory can be based (or not be based) on idealizing assumptions, it is more likely to happen in the case of ideal theories. Indeed, some critics of ideal theory believe that *all* ideal theory is idealised – but I have argued that this need not be the case.

4. What—if any—role is there for ideal theory?

Assuming that everything that I have argued so far is valid, we can now ask what the role of ideal theory is. I will ask three questions: (a) Is ideal theory necessary for non-ideal theory? (b) How useful is ideal theory for non-ideal theory? (c) How limited, then, is the overall role for ideal theory in the field of normative social justice analysis?

(a) Is ideal theory necessary for non-ideal theory?

I mentioned earlier that for Rawls (1971/1999: 8) “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].” Indeed, it is generally assumed among contemporary theorists of justice that ideal theory is
necessary for non-ideal theory (Phillips 1985). But is it really true that ideal theory is needed for non-ideal theory and justice-enhancing action?

In his recent paper ‘What Do We Want from a Theory of Justice?’, Amartya Sen (2006) denies that this is the case. He presents a strong critique of the ‘transcendental’ approach to justice, and advocates instead a ‘comparative’ approach to justice. If my reading is correct, than Sen’s definition of ‘transcendental’ is precisely the definition of ideal theory as laid out in section 2. Sen defines a ‘comparative’ approach to justice as an approach which allows us to say which of two social states is more just than the other. This is clearly an important part of non-ideal theory, since to judge the justice-enhancing degree of a policy or action, one needs to compare the expected corresponding social states.

Sen (2006) argues that many cases of blatant injustice do not require an ideal theory of justice before one can take justice-enhancing measures. We don’t need an ideal theory of justice to know that children who suffer severely malnutrition while there is enough food to feed everyone are treated unjustly. Sen argues that a complete ideal theory of justice is too demanding, and does not allow us to remain silent on some issues where we believe incompleteness is in fact the furthest we can go. In such cases we could start by acting on those aspects of injustice which are blatant: if we are living in social state A (undernourished children) and can choose between B1 (no justice-enhancing action but working on a complete ideal theory of justice) or B2 (working towards solving the problem of blatant injustices, such as hunger among children), then B2 is clearly a more just social state than B1 and A, and hence we know that moving from A to B2 is a justice-enhancing action. We don’t need to aspire to completeness at the level of ideal theory before being able to make confident claims about at least some justice-enhancing actions.

I believe that Sen’s arguments are correct, but they only show that ideal theory is not always necessary for non-ideal theory and policy design and implementation. I do think there are cases where ideal theory is necessary in order to advance justice. In cases where the gap between ideal justice and the actual situation is relatively small or not easily detectable by the untrained or uninformed eye, ideal theory may be necessary to tell us where the remaining justice gaps are. I believe this to be the case with gender injustice in contemporary post-industrial societies. Since many people believe that women and men are treated equally, and since many of the causal sources
of gender injustice are not visible to the (untrained) eye, the remaining gender injustices are subtle and, at any point in time, may be rather small.\footnote{Nevertheless, even if they are small they remain unjust, and small inequalities can accumulate into major inequalities over time (Valian 1998). The most recent time budget studies for Belgium show that full-time employed women work slightly longer every day than their male partners, but at an annual base this amounts to 4 full-time working weeks. Seen from this perspective, what may look like a trivial inequality on a daily basis may become significant on an annual basis.} We therefore need an ideal-theoretical (even if partial) account of gender justice in order to show where the remaining gaps are and then to decide what measures are needed to close these gaps (Robeyns 2007).

In sum, while I agree with Sen’s argument that ideal theory is not always necessary in order to decide which justice-enhancing measures to take (as in cases of gross injustices), I disagree with the underlying insinuation that ideal theory is never necessary for non-ideal purposes. The case of gender injustice in contemporary post-industrial societies shows that there are examples of ‘non-obvious injustices’ where ideal theory may serve as a test to show that there is still a remaining justice gap (which is a necessary step before one can proceed to action).

**b) Is ideal theory useful for non-ideal theory?**

Ideal theorists may grant that while ideal theory may not always be necessary for non-ideal theory, it is nevertheless very useful in the majority of cases. Is this true? To answer this question, I think it is necessary to distinguish between ideal theory based on strongly idealizing assumptions, and ideal theory that does not rely on such assumptions.

If ideal theory is based on idealizing assumptions, then it can in fact be rather misleading instead of being helpful. I think this was one of the things that was actually at work in Elizabeth Anderson (1999) much-discussed article ‘What is the point of equality?’. In this article, she shows the counterintuitive and often absurd consequences in the real world of what she named ‘luck egalitarianism’, i.e. liberal egalitarianism inspired by Dworkin and his choice/circumstance distinction (or ambition-sensitivity and endowment insensitivity principles). Following Dworkin, luck egalitarianism holds people responsible for those choices that are the result of their own ambitions, but argues that people should not be held responsible for any inequality resulting from their (unchosen) endowments. However, as I discussed
earlier in this paper, this only holds if certain background conditions are met, such as a fair distribution of resources, authentic preferences and the elimination of prejudice. These are idealising assumptions, and as far as our current knowledge on social and cognitive psychology goes, the superhumans that luck egalitarianism presupposes are beyond any realistic imagination. Thus, any policy recommendations that are derived from luck egalitarianism which also draw upon these idealizing assumptions, will therefore be flawed in the real world because these luck-egalitarian superhumans do not exist (nor does luck egalitarianism offers us any plausible account on how to re-educate ordinary people into being these kind of superhumans).

What about the usefulness of ideal theory that is not based on idealizing assumptions? I think we have enough convincing cases of ideal theoretical principles that, if implemented in practice, can be counterproductive or even lead to rather disastrous consequences. Take the case of minimal justice for children. Suppose (not implausibly), that ideal theory tells us that children should, as a matter of justice, be entitled to a decent education and should be protected from having to work before a certain age. What does this imply for justice in non-ideal circumstances? As Pierik and Houwerzijl (2006:194) point out, “rejecting child labor on moral terms is one thing: fighting it is quite another matter”. By analysing recent events, they showed that some measures inspired by the principles that children should not work, such as US legislation imposing a ban on the import of goods produced by children, have had rather disastrous consequences. Legislation of the countries where these children live often doesn’t make much of a difference – if any difference at all. Even if one accepts that in principle child labour is a gross violation of minimal justice for children, the most justice-enhancing measures both in the short-term and the long-term may nevertheless be based on an acceptance that in the short term child labour cannot everywhere be eliminated. But such attitudes and the design of policies and actions that would follow, are based on a kind of pragmatism that is at odds with ideal theory in general.

(c) The inherently limited role of ideal theory

If the analysis in this paper is correct, then it follows that ideal theory plays a rather limited role: it depicts the shape and the location of the lighthouse of justice, but it does not tell us how to get closer to the lighthouse from where we are now. That work
remains to be done by non-ideal theory, and action design and implementation. Many ideal theorists acknowledge this in their writings. Rawls stresses the limited role of ideal theory, when he writes that ideal theory is only worth of study because it is the fundamental part of the theory of justice and essential for the nonideal part; in addition, Rawls acknowledges that questions of non-ideal theory as “more pressing” (p. 9). Nevertheless, when discussing the tasks of political philosophy, he only includes ideal theory, without even suggesting who should then bear the task of developing the non-ideal theory of justice and the mixed theoretical-empirical work of actions design and implementation (Rawls 2001: 4-5; 2007; 10-11). For Rawls, political philosophers need not be concerned with non-ideal theory.

Much of contemporary post-Rawlsian theoretical work on justice has proceeded in this fashion: firstly, by assuming that ideal theory is justified since it would be fundamental (and hence necessary) for non-ideal theory; and second, by limiting their own work to ideal theory, often without telling the readers where to look for the non-ideal work which is needed to provide a solution to the genuine problems.

In the writings of ideal theorists non-ideal theory and action design may be or is explicitly considered important: but from their practice one may wonder whether this is not mere rhetoric and lip-service. 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 work, they do not think that symmetrical scholarly obligations rest on them to ask how useful their ideal theories are for real life problems or what needs to be done to make them really contribute to enhancing justice. Ideal theorist would do well to acknowledge in their scholarly practice these limitations of ideal theory – at any rate more than is currently the case.

---

9 See for example Mason who writes: “In order for someone to be able to work *in a sophisticated way* at the second and third [i.e. less abstract and more empirically informed] levels, he or she will need to understand the various moves that are made at the first level [i.e. the most abstract level where idealizing assumptions are allowed] (2004: 265-266, italics added). Mason thus suggests that all applied or non-ideal theory needs the most ideal (and perhaps idealized) of ideal theory in order to be decent (‘sophisticated’) work: to my mind this grossly overestimates the usefulness and the necessity of ideal (and certainly idealized ideal) theory for non-ideal theory. But if one truly believes this statement, then it explains why one can be contented spending a whole live on (idealized) ideal theory.
5. Concluding remarks

What conclusions can we draw from the arguments presented in this paper?

First, ideal theory and non-ideal theory are currently not well defined in the literature on theories of justice. It would help if we could somehow agree on what defines ideal theory and non-ideal theory.

Second, part of the lack of clarity is based on the fact that ideal theory is often defined based on one of the dominant characteristics of contemporary ideal theory in practice, namely idealizing ideal theory, that is, ideal theory constructed upon (strongly) idealizing assumptions. I have tried to show that (1) ideal theory need not be idealizing, but that some of the major contemporary theories of justice are idealizing indeed, and (2) that the role of ideal theories with idealizing assumptions that are constitutive for their principles of justice, is limited to playing an intellectual game: suggestions of any problem-solving relevance of such ideal theory are misleading.

Third, I have tried to argue that the role of non-idealizing ideal theory is in many cases limited. Ideal theory is sometimes not even necessary for non-ideal theory and policy and action design. Moreover, principles of ideal theory can be counter effective if one tries to implement them in the real world.

Fourth, in contrast to Sen, I do believe that some cases of ideal theory are relevant to the real world: not all ideal theory is simply an intellectual game.

Taken together, my reading of the literature on theories of social justice suggested a modest and rather limited role for ideal theory in thinking about justice. This may lead one to question Rawls’s view that political philosophy should not be bothered with non-ideal theory and the current status of ideal theory among political philosophers. In my view, political philosophers should stop their flawed practice of constructing theories based on idealizing assumptions altogether, and should shift their focus away from ideal to non-ideal theory.

References


