

On the Apparent Paradox of Ideal Theory*

LAURA VALENTINI

Political Science, University College London

Published in *The Journal of Political Philosophy*

It is often remarked that contemporary liberal egalitarian conceptions of justice are developed in ‘ideal theory’. In spite of its popularity, the notion of ideal theory tends to be employed somewhat loosely, to indicate any theory constructed under false, that is, idealised, assumptions, which make social reality appear significantly ‘simpler and better’ than it actually is.¹ Admittedly general, this definition accurately characterises most contemporary liberal egalitarian accounts of justice. Just to mention a couple of examples, John Rawls famously simplifies the problem of social justice by assuming that society is self-contained, populated with fully capable adults and exists under favourable natural and historical conditions.² He also improves the moral features of society by assuming full compliance on the part of the agents falling under the purview of his principles.³ Similarly, Ronald Dworkin’s theory of equality of resources is developed in what he calls ‘the ideal world’.⁴ That is, it assumes away both the technical limitations we encounter in actual societies, and the fact that human beings often lack the political will to follow principles of justice.

However popular, this idealising trend has become an object of concern among political theorists: there is a worry that the gap between liberal egalitarian ideal theories and our non-ideal circumstances might be unbridgeable. Principles designed assuming ideal conditions, so the worry goes, are ill-suited to guide action in the real world, where such conditions do not hold. Given that a capacity for guidance is widely considered a necessary attribute of any sound normative theory, the ‘guidance critique’ seems to pose a serious threat to liberal egalitarian thinking as a whole.⁵

Even more troubling, responding to this threat by giving up ideal theorising does not look like a viable option. There is indeed a sense in which resort to ideal theory is inescapable: How could we

formulate judgements about the justice and injustice of society, let alone promote institutional reform, without having an ideal of what a just society would look like? Moreover, how could such an ideal be constructed without making idealised assumptions? This being the case, liberal egalitarians apparently find themselves in the contradictory situation where their ideal theories are *unable* to guide action and, yet, at the same time, *indispensable* in guiding action. This, which I call ‘the paradox of ideal theory’, can be stated as follows:

- a) Any sound theory of justice is action-guiding.
- b) Any sound theory of justice is ideal.
- c) Any ideal theory fails to be action-guiding.

Is there a way out of this paradox? In this article I argue that there is. I reach this conclusion through a two-stage argument. In the first stage (sections I, II, III) I articulate and examine each of the claims contained in the paradox, with a view to clarifying the terms of the discussion and putting the target of my investigation into sharper focus. In section I, I distinguish between fact-sensitive and fact-insensitive approaches to justice. Since the latter reject claim (a), I restrict the scope of the investigation to the former, focusing on John Rawls and Ronald Dworkin as their most eminent proponents. In section II, I analyse claim (b) and, in section III, I spell out claim (c), namely the guidance critique, in greater detail.

In the second stage of the argument (sections IV, V, VI), I assess the validity of the guidance critique in relation to four prominent examples of ideal theory, and show that it successfully undermines only some of them. This suggests that there is nothing wrong with ideal theory *per se*, but rather that there can be good and bad forms of ideal theorising. In section VI, in particular, I explore this suggestion and argue that an ideal theory of justice becomes vulnerable to the guidance critique only if it *entails* an idealised account of the subject to which it is meant to *apply*, and not merely by virtue of it being constructed under idealised assumptions. This leads me to conclude that the paradox can be

overcome: liberal egalitarian theories of justice which aspire to be action-guiding (a) are indeed ideal (b), but only a subset of them succumbs to the guidance critique (c).

I. Claim (A): Justice and Guidance

A tendency to theorise ‘in the ideal’ is one of the hallmarks of contemporary liberal egalitarianism. For this reason, it is natural to suppose that the paradox of ideal theory affects liberal egalitarians generally. This, however, would be too hasty a conclusion. Whether a specific liberal egalitarian outlook has to confront the paradox depends on its endorsement of (a), namely the claim that any sound theory of justice is action-guiding. At first, such a claim appears unquestionable: How can a sound *normative* theory lack a capacity for guiding action? Surprising as this may be, some forms of liberal egalitarianism seem to be unconcerned with the issue of guidance, thereby falling outside the scope of the present discussion.

Before proceeding further, let me then refine the target of my investigation, and distinguish, following G.A. Cohen, between two types of theorising about justice, both of which are present in contemporary liberal egalitarian thinking: fact-sensitive and fact-insensitive.⁶ The main difference between these two approaches lies in the way they conceive of justice in relation to empirical facts, especially facts about human nature and its operation within existing social arrangements.

Typically, fact-sensitive theories focus on the justification of political institutions where participation is non-voluntary and individuals’ fundamental interests are at stake. They conceive of the question of justice as an inherently practical one, to do with the legitimate exercise of political power.⁷ For fact-sensitive theories, the question of justice only arises in the presence of a certain ‘social fact’: the existence of power-relations channelled through common institutions.⁸ Since principles of justice are meant to regulate the operation of such institutions, several factual considerations have to be taken into account when designing them.

Fact-insensitive theories, on the other hand, abstract away from existing practices, and attempt to define justice ‘in its pure form’, that is uncontaminated by facts. They think of justice as a fundamental moral value, characteristically expressed by some principle of distributive equality, the nature of which is *insensitive* to (indeed, has nothing to do with) existing facts, including the ‘social fact’ of political power.⁹

Even though it is tempting to regard fact-insensitive and fact-sensitive approaches to justice as lying on a continuum, from ‘fully ideal’ to ‘less ideal’, this would be a mistake.¹⁰ The difference between them is not just a matter of degree. These two approaches do not assume the same notion of justice, and then proceed to elaborate theories pitched at different levels of ‘idealism’. Instead, they conceive of the question ‘What is justice?’ in completely different terms:

1. Fact-insensitive theories ask: ‘What is justice as such, in its pure, fact-free form?’
2. Fact-sensitive theories ask: ‘What principles should govern the exercise of political power for it to be justified?’¹¹

While question (2) is premised on the idea that principles of justice should perform a certain function, namely that of governing the exercise of political power, question (1) has no such functional connotation. Consequently, while plausible candidate principles offered in answer to question (2) must display some capacity for guidance, no such guidance-requirement applies to question (1). As G.A. Cohen, the most radical proponent of the fact-insensitive approach, says, justice might ‘not [be] something that the state, or, indeed, any other agent, is in a position to deliver’.¹² Rather than telling us ‘what we should do’, political philosophy is about ‘what we should think, even when what we should think makes *no* practical difference’.¹³

In light of this, there is reason to leave fact-insensitive theorists out of the present discussion. Even though their highly abstract theorising naturally invites talk of ideal theory, they do not consider a

capacity for guidance as a necessary condition for the validity of a conception of justice, and are therefore unmoved by the paradox. Of course, this observation raises the deeper question of whether it makes sense to understand conceptions of justice as fact-insensitive theorists do.¹⁴ But this question clearly differs from the one I have set out to tackle in this article, to which I now turn.

Among the liberal egalitarian theorists who embrace a fact-sensitive outlook, John Rawls and Ronald Dworkin are probably the most celebrated as well as the most criticized.¹⁵ As for Rawls, his approach to justice is fully in line with the fact-sensitive paradigm I have described. First, Rawls understands principles of justice as criteria for the justification of the basic rules governing the exercise of political power in a contemporary liberal democracy. As he puts it, the question ‘justice as fairness’ attempts to answer is: ‘[I]n the light of what reasons and values – of what kind of a conception of justice – can citizens legitimately exercise ... coercive power over one another?’¹⁶ His principles of justice, then, express the reasons and values which should inform the morally appropriate distribution of resources within society and determine the structure of citizens’ power-relations. When such principles are violated, instead of a society of equals who *legitimately coerce* one another through participating in common institutions, we have a political system where some are *oppressed* by others.

Rawls’s conception of political philosophy is also inherently concerned with the practical.¹⁷ On his view, resorting to idealised conceptions of society and the person ‘is a way of continuing public discussion when shared understandings of lesser generality have broken down’.¹⁸ The task of political philosophy and ideal theory is to reduce (hopefully resolve) such disagreements,¹⁹ offering a ‘public framework of thought’²⁰ from within which to assess the justice or injustice of the current distribution of resources within society.²¹

Similarly, Ronald Dworkin understands justice in connection with the justification of political power. For him, justice defines the principles according to which governments, the *loci* of political power, should treat their citizens.²² Moreover, like Rawls, Dworkin understands the point of theorising about justice as an eminently practical one. In his view, political philosophy should ‘respond to politics’ and abstract philosophical argument should begin in real life ‘because only then it is likely to have the

right shape, not only *finally to help us*, but also finally to satisfy us that the problems we have followed into the clouds are, even intellectually, genuine not spurious'.²³

For both Rawls and Dworkin, then, the point of ideal theory is to guide action in non-ideal circumstances. This concludes my analysis of claim (a). As we have seen, not all liberal egalitarian theorists endorse this claim, hence not all of them are troubled by the paradox. Those who endorse it, most famously John Rawls and Ronald Dworkin, think of principles of justice as offering criteria for the justification of the exercise of political power, and believe that a theory of justice should be action-guiding in the sense of offering a framework of thought from within which to assess, criticize and reform the way power is exercised within society. With these ideas clearer in mind, we can move on to claim (b), and take a closer look at the notion of an ideal theory, which I have so far left admittedly vague.

II. Claim (B): Are Theories of Justice Ideal?

The second claim within our paradox states that any sound theory of justice is ideal. Is this a correct assertion? To answer this question we need to investigate the notion of an ideal theory in greater detail. As I said, this notion tends to be employed somewhat loosely in contemporary political philosophy. Close scrutiny reveals that it can take at least two different meanings.²⁴ First, a theory may be 'ideal' in a non-technical sense, insofar as it proposes an ideal of a fully just world towards which we should aim. Intuitively, it would seem that, unless a theory is ideal in this way, it cannot qualify as a normative theory of justice.²⁵ Second, and more interestingly, a theory can be ideal in the technical sense of being designed under idealised, i.e., false, assumptions.²⁶

The relation between these two understandings of 'ideal theory' is unclear. The question of whether a theory can be ideal (in a non-technical sense), without resorting to dramatic idealisations has generated considerable debate in the literature, and engaging with it would take me too far from the

present discussion.²⁷ In what follows, I will therefore set aside the issue of whether normative ideals can be designed in the absence of idealisations, and limit my analysis to those theories which are ideal in *both* senses. Since most contemporary liberal egalitarian theories of justice are ‘doubly ideal’ in this way, and since these are the targets of the guidance critique, this restriction does not significantly affect the scope of the article.

For instance, both Rawls’s and Dworkin’s theories describe ideals of fully just societies and are elaborated under false (idealised) assumptions. Central to the justification of Rawls’s theory is a hypothetical agreement situation where citizens’ representatives are asked to choose the principles of justice which will govern the basic institutions of society. The parties to the agreement and the conditions under which they deliberate are deeply idealised. In particular, the parties are assumed to be indifferent to one another’s fate and to deliberate under a ‘veil of ignorance’, which prevents them from knowing the real-world identities of those they represent.²⁸ They are only aware of ‘the general facts about human society’ – such as, facts about economics and psychology²⁹ – and deliberate assuming that the people they represent have full mental and physical abilities and will comply with the demands of justice.³⁰

The outcome of this hypothetical choice situation is Rawls’s famous two principles: the equal liberty principle and the difference principle (with a fair equality of opportunity constraint). The former secures citizens’ basic civil and political liberties, the latter a form of distributive equality among them. The point here is that none of the conditions under which these principles are derived can be said to reflect the characteristics of imperfect, real-world, human beings.

Dworkin’s theory of justice is similarly constructed under idealised assumptions, in what he calls ‘the ideal ideal world of fantasy’.³¹ This is an imaginary scenario where a group of shipwreck survivors stranded on a desert island deliberate about how to achieve a fair distribution of resources among them.³² They resort to a two-stage procedure, involving, first, an auction of the goods present on the island, where all participants have the same purchasing power, and everyone bids until no one envies the bundle of resources allocated to others.³³ In addition, to address the unfair disadvantage suffered by

those with poor natural endowments (the untalented and the disabled) the auction is supplemented by a hypothetical insurance scheme. Participants are asked how much they would be willing to pay to insure against natural disadvantages, and the insurance they would buy is then employed to address the inequalities suffered by the untalented and the disabled.³⁴

Finally, but no less importantly, Dworkin's survivors are thought to deliberate against a fair background, establishing what they will be free to do and forbidden from doing with the acquired resources, and assuming away the effects of prejudices and stereotypes on each one's final share.³⁵ For instance, no one should use resources to threaten others' security, and no one should end up being worse off due to others' prejudices. The principles derived through this complex thought experiment, Dworkin says, should be further constrained by considerations of technical feasibility, and it is these less idealised principles that should then be used for assessing the distributive performance of existing societies.³⁶

Interestingly, both Rawls and Dworkin construct principles of justice assuming away any serious failure to comply with what 'humanly achievable justice' requires.³⁷ If they took into account agents' lack of motivation to follow the commands of justice, they would hardly obtain principles of *justice* at all. If motivational and behavioural weaknesses (beyond a reasonable extent) were factored into the elaboration of a theory of justice, such a theory would lack the necessary critical distance to assess the *status quo*, and would thus be seriously defective.³⁸ Having said that, the question before us is whether this and other idealisations necessarily rob ideal theories of any practical import, as maintained by the guidance critique. To answer this question, I now turn to claim (c), and explore the contents of such a critique more in detail.

III. Claim (C): The Guidance Critique and its Interpretations

As I said at the outset, ideal theorising has been repeatedly attacked on the grounds that it fails to guide action in the real world. This admittedly vague formulation of the guidance critique can be interpreted in at least three different ways (GC1, GC2, GC3). The first two, I argue, do not raise interesting theoretical challenges, whereas the third does.

GC1: It might be argued that ideal theory lacks a capacity for guidance because it fails to *motivate* existing agents: in practice, ‘it does not work’. On this view, a conception of justice fails to be action-guiding when, as a matter of fact, people do not follow it.³⁹ The fact that people lack the motivation to act in accordance with principles of justice is certainly regrettable. But, as I noticed at the end of the previous section, it hardly counts as a reason against the validity of a conception of justice. The point of a theory of justice is precisely to give us a conceptual framework from within which to criticise existing agents who do not conform with it. So long as, given our best account of human motivation, it is *reasonable* to expect compliance, the fact of actual non-compliance tells us nothing about the adequacy of the theory itself.⁴⁰

GC2: Ideal theories of justice might be said to lack a capacity for guidance because they are not *immediately* applicable to day-to-day political decisions. It is indeed undeniable that they leave answers to questions such as how citizens and officials should act in this or that particular situation largely indeterminate. Is this a serious charge against ideal theorising? I doubt it, for such a charge presupposes unreasonable expectations as to what any plausible normative theory can offer. As Kant himself notes, ‘no matter how complete the theory may be, a middle term is required between theory and practice, providing a link and a transition from one to the other’.⁴¹ That is, in order to become practically relevant, a theory needs to be brought to bear on specific cases through judgement, which involves a careful assessment of the facts of the matter.⁴²

As suggested by Onora O’Neill, the role of principles is to help us single out the relevant facts and evaluate their implications for action – implications which cannot be already contained in the principles themselves.⁴³ Again in Kant’s words, principles ‘are abstracted from numerous conditions which, nonetheless, necessarily influence their practical application’.⁴⁴ Moreover, the availability of

common principles is not itself a guarantee of convergence in judgement: there is always the possibility for principles to be given different interpretations. Such a possibility, however, does not deprive of worth the enterprise of looking for principles which offer a basis for mutual justification; recall Rawls's observations on the need to construct a 'public framework of thought'. To deny the value of principles in this way is to deny the value of theorising as such, but this cannot be what proponents of the guidance critique have in mind.⁴⁵

GC3: The third formulation of the guidance critique amounts to the twofold contention, variously articulated by a number of critics, that ideal theory 'is at best morally irrelevant, at worst morally destructive'.⁴⁶ This charge can be summarised as follows: if we apply principles developed under ideal conditions to real-world circumstances – namely those circumstances for which they have allegedly been designed – we are bound to obtain morally counterintuitive results.⁴⁷ For instance, it would be self-destructive for a person to act in accordance with the principle 'you ought to be truthful' if she lived in a world where everyone else was dishonest. In such conditions, the principle loses its otherwise undisputed moral plausibility. Likewise, it would be morally absurd for one to always keep one's promises in a world where nobody else, or very few, always kept theirs.⁴⁸ As O'Neill points out, '[p]ractical reasoning that assumes that "ideal" predicates are satisfied will not reach conclusions safely and soundly for cases where they are not satisfied'.⁴⁹ Similarly, Annette Baier protests, 'descriptions of ideal societies imply *nothing* about how to act in this society'.⁵⁰

Notice that the claim is not that ideal theory is *insufficient* to guide action. This would simply reiterate GC2 which, as I said, is not problematic. Rather, the claim is that it is of no use.⁵¹ Put metaphorically, the problem is not that there is a gap to be bridged between ideal theory and non-ideal circumstances; the problem is that the gap is *unbridgeable*. There is no way, the critique says, in which ideal theory can be brought to bear on non-ideal questions. Referring to the specific non-ideal case of racial discrimination, for instance, Thomas McCarthy argues:

there are no theoretical means at hand for bridging the gap between a color-blind ideal theory and a color-coded political reality, for the approach of ideal theory provides no theoretical mediation *between* the ideal and the real – or rather, what mediation it does provide is usually only tacit and always drastically restricted.⁵²

Proponents of this line of criticism sometimes take it even further, arguing that attempts to bridge the gap by bringing the theory to bear on existing practice are most likely to be counterproductive. For instance, Charles Mills warns about the risk of ideal theorising being *ideological*. By focusing on the design of a fully just social order, he argues, ideal theories obscure, rather than illuminate, unjust power-relations, such as those based on gender and race. By assuming full compliance and enhanced human capacities, so the claim goes, ideal theories become blind to past and present injustices and, instead of offering conceptual tools for criticising them, they surreptitiously contribute to their perpetuation.⁵³

This third interpretation of the guidance critique is certainly more troublesome than the previous two. If ideal theorising were doomed to be irrelevant or counterproductive, then political philosophers broadly in sympathy with a Rawlsian approach would have a lot to worry about.⁵⁴ Do they? My answer is a moderate ‘yes and no’. To substantiate it, I briefly discuss GC3 in relation to several prominent examples of supposedly misleading idealisation. In particular, I consider how Rawls’s and Dworkin’s theories of justice deal with racial and gender discrimination, how Rawls’s ‘Law of Peoples’ approaches the question of global poverty, and how Will Kymlicka’s theory of minority rights handles the non-ideal case of illiberal national minorities. I argue that, while GC3 fails in the former two cases, it succeeds in the latter. The lesson ideal theorists should learn from this will be discussed in the final section of the article.

IV. Ideal Theory, Racial and Gender Discrimination

Questions of racial and sexual discrimination are well-known both for their actual political urgency and for falling outside the scope of ideal theories. Rawls acknowledges this omission, but also points out (in line with my earlier response to GC2) that ‘an omission is not as such a fault, either in [a] work’s agenda or in its conception of justice. Whether fault there be depends on how well that conception articulates the political values necessary to deal with these questions’.⁵⁵

GC3, however, claims precisely that ideal theory lacks the resources to do so. Its omissions are not harmless: by neglecting to discuss racial and sexual discrimination, ideal theory loses its capacity to guide action in the real world, where instead of memories from a morally flawed past, discrimination and oppression are facts of the present. Is this claim warranted? Focusing on the cases of gender and racial inequality in relation to Rawls’s and Dworkin’s theories of justice, I argue that it is not.⁵⁶ Let me begin with Rawls’s treatment of gender discrimination.

The issue of the relation of Rawls’s theory to gender injustice has been widely discussed by feminist theorists and particularly by Susan Okin.⁵⁷ She concedes that, even though Rawls does not directly discuss gender injustice, his theory contains valuable resources to address it. In *Justice as Fairness: A Restatement*, Rawls briefly takes up the issue of the justice of the family and recognises that, since the family is the fundamental unit of reproduction of society, it is to be regarded as a crucial constituent of the basic structure of society. His two principles apply to the family in an indirect fashion, setting *constraints* on the forms this institution may legitimately take, ruling out those which would violate the freedom and equality of citizens.⁵⁸ That is, if a certain way of arranging the internal life of the family were to result in an infringement of a citizen’s equal rights and opportunities, political measures (such as laws, incentives, creation of new infrastructures, etc.) to correct this infringement would be justified, in fact mandated, by the theory.

Following Okin's proposal, Rawls suggests that if a major cause of women's social disadvantage turned out to be the lack of material recognition for their contribution to society's continued existence through raising children and caring for the family, their contribution should be officially recognized, for instance, by promulgating laws entitling them to half of their husbands' earnings during the marriage.⁵⁹ From society's viewpoint, by taking care of children, and ensuring that they become fully capable citizens, women's activity is just as valuable as that of their partners and therefore deserves recognition.

What the basic structure 'should do' about this type of inequality – i.e., what measures would be most adequate – is something that cannot be determined without referring to the specific social contexts in question.⁶⁰ Whatever will best promote the end of social justice, understood as the absence of oppression between different social groups, be they women or blacks, requires all-things-considered judgements which cannot be deduced from the theory itself.⁶¹ Such judgements surely involve a great deal of intellectual effort, reliable data as well as political sensitivity, especially when injustices are not 'incorporated into recognized practice, if not the letter, of social arrangements'⁶² and therefore require 'an informed examination of institutional effects'⁶³ in order to be detected and remedied. However, the fact that applying a theory of justice to a certain social context is a complex intellectual operation does not entail that the theory is irrelevant to that social context.

Of course, it might be objected that 'the basic structure' can only do so much in remedying gender discrimination. Legal measures cannot directly address some of the deepest causes of women's disadvantage, especially when they belong to the private sphere. Even the best laws cannot eradicate all prejudices, stereotypes and clichés.⁶⁴ This may be true. Laws cannot *guarantee* the end of discrimination. Nonetheless, we should also bear in mind that institutions are powerful means by which to shape the social ethos. By ensuring that women enjoy equal standing in their capacity as citizens – through different context-sensitive measures – just social institutions can help foster a culture of respect between genders. Until relatively recently, arbitrary discrimination used to be embodied in the letter of the law. One may thus hope that, with the passing of time, changes in institutional ethos will contribute

to changes in private ethos, thereby reducing those forms of gender injustice which fall outside the ‘formal’ reach of the law.⁶⁵

At this point our objector might protest that this is too optimistic an estimate and that subtle forms of discrimination are bound to slip through the ‘formalistic’ normative machinery of ideal theories. Even if this were true, it would hardly count as a ‘guidance objection’ to ideal theory. Instead of signalling a problem of guidance, it points to one of scope, challenging the public focus of a political theory of justice. This is a legitimate critique against liberal egalitarianism, however it differs from the one which constitutes the object of my discussion. The former challenges the eminently political scope of contemporary theories of justice, while the latter focuses on their (in)capacity to redress relevant injustices *within* the political domain. A theory of social justice is, indeed, a theory of *social* justice: its aim and ambitions are limited to a particular normative sphere. It is therefore unreasonable to expect such a theory to address all forms of injustice occurring in both public and private relations. So long as the theory’s principles are plausible tools with which to secure citizens’ status as free and equal, the theory succeeds in achieving its aim.

To sum up, Rawls’s discussion of gender injustice in ideal theory may certainly be said to be incomplete (indeed, how could it be otherwise?), but it is not irrelevant or counterproductive. Similar considerations apply to Dworkin’s theory.

Like Rawls, Dworkin adopts what Ingrid Robeyns and Roland Pierik call a ‘*tabula rasa* approach’ to justice. He imagines his shipwreck survivors developing the ideal of a just distribution against a fair baseline, where all prejudices, stereotypes and other forms of injustice are assumed away.⁶⁶ Due to these idealisations, Dworkin’s theory has been accused of blindness towards socially generated inequalities, such as those connected with gender and race.⁶⁷ Is this accusation warranted? I believe not. The fact that socially generated inequalities are assumed away in the design of a theory of justice does not mean that the theory is, in turn, incapable of condemning such inequalities as unjust. Quite the opposite.

A just distribution has to be designed against a ‘fair baseline’,⁶⁸ which involves, among other things, assuming away racial and gender discrimination and its causes. If gender and racial stereotypes

were to be factored into the design of a just distribution, the result would hardly be a just distribution at all.⁶⁹ By assuming socially generated inequalities away, the theory enables us to appreciate that a just distribution must not be affected by prejudices and stereotypes.⁷⁰

Once again, the question of how to reach racial justice in non-ideal circumstances is a complex one. The answer cannot be determined a-contextually, independently of the particular circumstances to which it is meant to apply. For instance, if deep stereotypes still exist in societies where, formally, discrimination has been outlawed, affirmative action policies might be, all things considered, justified.

As Dworkin's discussion of affirmative action in *Sovereign Virtue* shows, when egalitarian principles have to be applied at the level of policy, both 'deontological' considerations of fairness and 'consequentialist' considerations of effectiveness have to be taken into account.⁷¹ If affirmative action were likely to increase racial prejudices and hatred, then alternative measures would have to be implemented to redress this kind of social disadvantage. Once again, the particular form of such measures – e.g., subsidies for the education of formerly discriminated against minorities, anti-racial discrimination campaigns and so on⁷² – cannot be directly derived from ideal theory itself, but can only be established on a case-by-case basis. This task certainly necessitates considerable theoretical work, and as critics rightly point out, ideal theorists largely neglect it.⁷³ But, as Rawls says, 'an omission is not as such a fault'.

The foregoing, admittedly sketchy, discussion has tried to show that Rawls's 'justice as fairness' and Dworkin's 'equality of resources' offer plausible conceptual apparatuses from within which to capture and possibly remedy the forms of gender and racial inequality marking contemporary liberal democracies. Of course, I am not suggesting that Rawls and Dworkin might not be wrong on some other counts. What I am scrutinising here is not whether their ideal conceptions of justice are sound, all things considered, but rather whether they are *doomed to be unsound* because, like *any* form of ideal theorising, they must succumb to GC3. In this section, I have attempted to show that this is not the case. Should we then conclude that there is nothing to GC3? That ideal theorising is freed from the risk of being irrelevant or ideological? Not quite. Interestingly, the guidance critique seems to have bite on

Rawls's approach to international distributive justice in *The Law of Peoples* and on Kymlicka's account of minority rights, to which I now turn.

V. Ideal Theory, Global Poverty and Minority Rights

The Law of Peoples presents a 'conception of right and justice that applies to the principles and norms of international law and practice', and particularly to the '*foreign policy* of a reasonably just *liberal* people'.⁷⁴ Instead of being concerned with the inward dimension of liberal societies' exercise of political power, the book focuses on its outward expression,⁷⁵ and offers a list of eight principles for assessing the legitimacy of its exercise.⁷⁶ Taken together, these principles cover both ideal (full-compliance) theory, which concerns well-ordered societies' mutual relationships, and non-ideal (partial compliance) theory, which concerns peoples' conduct towards non-well-ordered societies. These are societies that do not meet Rawls's standards because they are internally oppressive or externally aggressive 'outlaw states', or otherwise disadvantaged 'burdened societies'.⁷⁷

To evaluate the force of GC3, we must ask: Is *The Law of Peoples* irrelevant or misleading when it comes to the assessment of the conduct of actual liberal societies, and the ways they exercise political power in the international realm? Many critics have answered positively, arguing that Rawls's theory ignores some of the most pressing issues of international morality and therefore provides little guidance in real politics. To mention just a few of these omissions, *The Law of Peoples*, it is said, completely overlooks the morality of secession,⁷⁸ is silent about global environmental justice and the regulation of transnational trade,⁷⁹ and devotes no space to migration policies. However, it should by now be clear that, by themselves, such omissions cannot suffice to discredit the plausibility of Rawls's theory – as we know, GC2 is correct but unproblematic. Rather, the question is whether despite such omissions *The Law of Peoples* offers a plausible theoretical framework for reasoning about what *existing* liberal societies ought to do in order to exercise their international political power legitimately, and whether they actually do so.

Interestingly, Rawls himself takes a step in this direction, devoting part of his book to the discussion of principles 7 and 8, which are explicitly designed for non-ideal circumstances (the former sets constraints on the conduct of war, the latter establishes a duty of assistance between peoples). However, when discussing such principles, he only partially suspends the assumption of full-compliance, and asks how *just* – or as he puts it ‘relatively well-ordered’⁸⁰ – liberal societies should act towards either unjust outlaw states, or unfortunate burdened societies.

If the world were inhabited by well-ordered liberal societies, Rawls’s move would be unobjectionable. But this is not so. Things change when we replace his liberal peoples with actual liberal societies and ask how *they* should act towards poor societies and aggressive states. Consider, for instance, the plausibility of the duty of assistance liberals have towards burdened societies (principle 8). Rawls’s approach to international distributive justice distorts the reality of contemporary international relations. As forcefully argued by Thomas Pogge, to see this, one only needs to think about the perverse incentives generated by internationally recognised rules (which contemporary liberal societies also support), such as the international borrowing and resource privileges. By entitling whoever effectively controls a country’s territory – i.e., independently of their legitimacy – to borrow and use resources in the country’s name, such rules incentivise corrupt elites and officials to seize power in already disadvantaged societies, rendering the goal of development ever more remote.⁸¹ Similar considerations often apply to WTO negotiations. In real-world international politics, liberal societies take advantage of their superior bargaining power and negotiate ‘terms of cooperation’ that are particularly ‘burdensome’ to less powerful nations – terms from which such nations cannot realistically subtract themselves.⁸²

In light of this, as Pogge notes, urging existing liberal states to *assist* burdened societies while they have contributed and still contribute to the perpetuation of their precarious political and economic conditions, is a mischaracterisation of the moral relationship in question:⁸³ such a description does not apply, i.e., it is, technically, *irrelevant* to existing circumstances, which is precisely what GC3 claims. What is worse, the *ideology* component of GC3 is also correct in this case. Even if contemporary liberal

societies are far from well-ordered, *The Law of Peoples* generates the impression that they are, which obscures, rather than illuminates, the injustices for which they are responsible. It is not that they merely ‘fail to assist’ burdened societies, they actively place part of the burdens on them.⁸⁴ In short, once we attempt to bridge the gap between Rawls’s ideal theory and ‘real-world’ circumstances, we obtain misleading recommendations.

A similar gap between ‘the ideal’ and ‘the real’ may also be found in Kymlicka’s theory of minority rights. Before focusing on the guidance-flaws of such a theory, let me briefly outline its main claims. Kymlicka famously argues that participation in a societal culture – a territorially bound, institutionally developed and linguistically distinct community – is a precondition for a person to exercise her capacity for autonomy.⁸⁵ This is because societal cultures provide their members ‘with a wide range of choices about how to lead their lives’, choices which would be inaccessible outside *one’s own* societal culture.⁸⁶ In light of this, Kymlicka develops a Rawlsian liberal approach to minority rights, where access to one’s own societal culture is regarded as a ‘primary good’. Like any such good, access to one’s societal culture is a necessary condition for autonomy and should thus be fairly distributed.⁸⁷

Fairness in the allocation of access to one’s societal culture, Kymlicka says, can be guaranteed so long as state policies follow this principle: ‘*freedom within* the minority group and *equality between* the minority and majority groups’.⁸⁸ On his view, special rights should be denied to national minorities when their exercise would violate the basic civil and political liberties of their members, but accorded if this is necessary to protect their societal cultures from the pressures of the wider society. That is, rights should be allocated with a view to preventing oppression both within and between minorities.

Since Kymlicka’s theory follows Rawls’s liberal paradigm, its status as an ideal theory should come as no surprise. What is surprising is that, unlike Rawls’s domestic theory of justice, Kymlicka’s is vulnerable to the guidance critique. Interestingly, the presence of a gap between his ideal theory and our non-ideal circumstances is signalled by Kymlicka himself, and particularly by his puzzlingly tolerant attitude towards real-world cases involving illiberal minorities.⁸⁹ I say ‘puzzlingly tolerant’ because it would seem to follow from Kymlicka’s theory that illiberal minorities should be forced to become

liberal. Yet, Kymlicka argues that a liberal state has no ‘authority’ to force a self-governing illiberal national minority to abide by liberal principles. Unless the minority is perpetrating gross violations of human rights, torturing or enslaving its members, it cannot be coerced to conform with liberal principles.⁹⁰

This apparent inconsistency in Kymlicka’s thought has generated much debate in the literature.⁹¹ Here I propose to look at it through the lens of the ‘guidance critique’, as revelatory of a significant gap between Kymlicka’s theory and his real-world prescriptions.⁹² When it comes to real-world guidance, one might say, Kymlicka refuses to ‘apply’ his theory, and develops prescriptions starting from scratch, working bottom-up rather than top-down. In so doing, he seems to acknowledge that his own theory is *irrelevant* to existing injustices. And this is because, even if within illiberal minorities people’s rights are violated, it is unclear ‘what third party (if any) has the authority to intervene’.⁹³

In real-world circumstances, not only national minorities, but also liberal states, are *not* perfectly just as Kymlicka’s theory assumes. In other words, the theory considers how *just* liberal states should act towards *just* national minorities,⁹⁴ and remains silent about real-life cases where minorities are illiberal and liberal states far from just. Unlike the international Rawls, who refuses to drop the assumption that liberal societies are ‘well-ordered’ in non-ideal theory, Kymlicka drops ideal theory altogether when it comes to the moral assessment of non-ideal circumstances. True, as Kymlicka says, and as we have seen when discussing GC2, there is a difference between the identification of a liberal theory of minority rights and its application – or ‘imposition’ as he puts it.⁹⁵ However, Kymlicka’s real-world prescriptions can hardly be regarded as an application (or imposition) of his own theory.

In light of these examples, supporters of GC3 seem to have some point: there are cases in which ideal theory has *no* relevance to real-world circumstances. Nonetheless, given the conclusions of the preceding section, we can reject the claim that ideal theories are *never* action-guiding. Instead, we are left with the following puzzle. All ideal theories contain numerous idealisations, but only some of them succumb to the guidance critique: Why is that?

VI. Distinguishing Between Good and Bad Ideal Theories

To make sense of the disanalogies between the theories examined so far, we need to test the tenability of their idealised assumptions against their respective aims, namely the justification of the exercise of domestic and international (coercive) political power.

From this perspective, it becomes clear that the idealisations contained in Rawls's and Dworkin's domestic theories of justice do not prevent them from incorporating an empirically plausible account of the channels through which power is exercised within society. Rawls does this through the notion of 'the basic structure', the complex set of institutions and public rules that govern the allocation of benefits and burdens among citizens. Though less explicitly, Dworkin does the same by focusing on society's legal system and government, which are precisely the means by which the most grievous forms of social injustice can be addressed. Neither in Rawls's nor in Dworkin's case does the *subject* of justice end up being idealized.⁹⁶

Things look different when we turn to *The Law of Peoples* and Kymlicka's *Multicultural Citizenship*. As for Rawls's international theory, this asks how almost self-contained political communities ought to behave towards one another, on the assumption that, since they are well-ordered, a just (fair) background is in place.⁹⁷ But because real-world societies are not well-ordered, such a background does not exist, and it is precisely its absence that gives rise to the question of international justice in the first place. (This point is also noted and discussed by Simon Caney and Thomas Pogge.)⁹⁸

In ideal theory, where liberal societies are well-ordered, there is an implicit assumption that their trading relations and agreements are fair, that none of them would exploit their superior wealth and bargaining power to make others agree on terms unfavourable to them.⁹⁹ This is a highly desirable ideal, but too far from the status quo. This is why the 'transition from ideal theory to non-ideal circumstances'¹⁰⁰ cannot be successfully carried through in the case of *The Law of Peoples*. A successful transition would require abandoning the assumptions (idealizations) that liberal societies are well-

ordered and almost self-sufficient, which are among the basic ‘building blocks’ of Rawls’s theory, giving shape to its framework for looking at the world.

As a result, instead of helping us make sense of what international justice requires today, such a framework proposes what Allen Buchanan calls a set of ‘rules for a vanished Westphalian world’.¹⁰¹ The problem is that, to try and ‘update’ the Law of Peoples by eliminating those idealised assumptions is to construct an altogether different theory.

This, as we have seen, is what happens with Kymlicka’s theory of minority rights. When we move from ideal to non-ideal circumstances, the theory plays no role in the design of moral prescriptions. Kymlicka’s theory tells us how just liberal states should treat just liberal minorities. But when it comes to unjust liberal states and minorities, the theory is silent, and moral prescriptions have to be constructed from scratch. In short, while *A Theory of Justice* and *Sovereign Virtue* design principles telling us what a just background ought to be like, *Multicultural Citizenship* and *The Law of Peoples* design their principles surreptitiously assuming that such a background is in place.

While suggesting that there is nothing wrong with ideal theory *per se*,¹⁰² the disanalogies between the theories under discussion also imply that the guidance critique is not without merit. Those who criticise ideal theory on this ground can be seen as pointing to ‘badly used’ idealisations, such as those found in *The Law of Peoples* and *Multicultural Citizenship*. Such idealisations build into a normative theory a false, i.e., idealised, account of the social phenomena the theory itself aims to put under moral scrutiny, in this way severely undermining its potential for guiding action in the real world.

The latter observation suggests a principled way of distinguishing between ‘good’ and ‘bad’ uses of idealisations. As we have said, all idealisations are false statements, which make the world appear simpler or better than it actually is. Our discussion in sections IV and V showed that good ones do so only ‘temporarily’, at the stage of theory construction, while bad ones do so ‘permanently’, as they are an integral part of a theory’s fundamental principles.¹⁰³ If we go back to the example of racial and gender discrimination, what makes ‘justice as fairness’ and ‘equality of resources’ successful against the guidance critique is that those facts about discrimination which are denied at the stage of theory

construction can be taken into account at the level of its application. In other words, Rawls's and Dworkin's principles of domestic justice do not assume but prescribe the absence of such forms of discrimination.

Similar considerations can be advanced in relation to other Rawlsian and Dworkinian idealisations, such as perfect rationality, mutual disinterestedness, ignorance about one's self and preference authenticity. Since they are introduced at the level of the original position and 'desert island' thought experiments, such assumptions are part of the theories' overall 'devices of representation' (as Rawls would say) which are not meant to model existing human conditions. What these thought experiments articulate are the circumstances *under which* it seems plausible to construct a theory of justice, not the subjects or agents *to which* a theory of justice should apply. Rawls's and Dworkin's idealisations would indeed prevent their theories from being meaningfully action-guiding if they were somehow entailed by their principles, but they are not.¹⁰⁴ In other words, the theories would fail to be action-guiding *due* to their idealisations *if* fulfilling their principles required citizens to be fully rational, their preferences to be independent and unbiased by prejudice and so forth. But this is not the case: such idealisations are not part of Rawls's and Dworkin's principles, they are merely part of the arguments supporting them.

O'Neill is thus right in arguing that '[p]ractical reasoning that assumes that "ideal" predicates are satisfied will not reach conclusions safely and soundly for cases where they are not satisfied'.¹⁰⁵ What I have shown here, is that the appeal to ideal predicates in the *design* of a theory of justice does not inevitably lead to unsound practical conclusions, at least so long as such predicates are not assumed *as* a condition for the *applicability* of the theory.

In *Multicultural Citizenship* and *The Law of Peoples*, by contrast, the facts about the real world that Rawls and Kymlicka deny when constructing their theories cannot be re-introduced at the level of their application while leaving the theories intact. Such facts are not 'innocently' denied at the level of theory construction, but their denial is also presupposed as a condition for the applicability of the theories themselves. The idealisations built into Rawls's international original position are treated as facts about

the agents and circumstances to which his ‘Law of Peoples’ is meant to apply.¹⁰⁶ Were such facts reintroduced in Rawls’s theory, it would no longer be the theory that it is. As we have seen, an example of this phenomenon is offered by Kymlicka’s discussion of how real-world liberal states ought to treat illiberal minorities. Such a discussion falls outside the purview of his ‘ideal theory’ of minority rights precisely because it is framed within a context where the idealisations which constitute a condition for the applicability of the theory – i.e., that liberal societies and minorities are ‘just and legitimate’ – no longer hold. In short, the agents to which Rawls’s and Kymlicka’s theories are meant to apply, namely just liberal societies, do not exist, and this is why these theories are irrelevant, if not misleading, when applied to real-world circumstances.

Arguably, the validity of these reflections about good and bad idealisations is not unique to normative theorising, but extends to non-normative fields of enquiry as well.¹⁰⁷ Consider the following two examples drawn, respectively, from physics and economics.¹⁰⁸ Famously, Galileo’s Law of Falling Bodies is elaborated under the idealised assumption that friction does not exist. In spite of it being ‘false’ of existing falling bodies, such an idealised assumption is widely regarded as appropriate for theoretical purposes. Galileo’s Law is not charged with being unsound because it is arrived at by assuming friction away. Such an assumption does not make Galileo’s Law ‘insensitive to friction’, so to speak. In fact, friction can be taken back into account when such a Law is employed to make predictions about falling bodies in the real world.

Now consider assumptions often made in standard economic theory, such as the assumptions that individuals are perfectly rational, fully informed and completely independent choosers.¹⁰⁹ Unlike Galileo’s, these assumptions have been repeatedly accused of being misleading, giving rise to models which misrepresent the phenomena they intend to describe. If we were to apply the theory, taking into account the motivational complexity of existing human beings, we would (arguably) end up repudiating it altogether.¹¹⁰

Parallel to the cases of ideal normative theory explored in this article, while the examples from physics and economics both contain idealisations, these are sound in the first example, but unsound in

the second. And again parallel to such cases, this discrepancy can be explained by noting that while in the former example the relevant facts can be reintroduced at the level of the theory's application – i.e., the theory is 'sensitive' to them – they cannot in the latter.

Before concluding it is important to note that these are just examples. The distinction between good and bad ideal theories is not a clear-cut one. Ideal theories are not either good or bad: they are better or worse. For instance, if the conditions of applicability of a theory are slightly idealised with respect to real-world circumstances, the theory might still retain its normative validity. Moreover, establishing whether an ideal theory is vulnerable to the guidance critique requires sustained argumentation. Apart from the most extreme cases – e.g., theories designed for angels – examples are bound to be contestable, and those offered here are no exception. This, however, does not affect the validity of the article's conclusion, which I am about to summarize.

VII. Conclusion

I started my discussion with what I have called the paradox of ideal theory, namely the claim that ideal theory is both necessary for guidance, and yet incapable of offering guidance. After some conceptual clarification, I have evaluated the force of the guidance critique (in its third version, GC3) by analysing four examples drawn from ideal theories which all aspire to be action-guiding. Interestingly, the four cases have yielded different conclusions, showing that there is nothing wrong with ideal theorising *per se*, but that there can be good and bad forms of it. Whether an ideal theory is good or bad depends on whether it *entails* a false, idealised, account of the *subject* to which it is meant to apply.

The paradox can thus be overcome: liberal egalitarian theories of justice which aspire to be action-guiding (a) are indeed elaborated in ideal theory (b), *but* not all of them are vulnerable to the guidance critique (c). The mere presence of idealisations in the assumptions under which a normative theory is constructed is insufficient to establish that the theory is unsound *because* it fails to guide action.

That said, even if the guidance critique does not provide a decisive objection against ideal theorising, it has the merit of reminding theorists who subscribe to a fact-sensitive approach not to lose sight of those *facts* that are *relevant* to the question they are attempting to answer. Keeping the facts in sight does not entail subservience to the status quo; rather, it is crucial for getting the distance between the ideal and the real right, so as to produce a theory that is both critical and action-guiding.

Notes

* This article was originally prepared for the ECPR Workshop ‘Social Justice: Ideal Theory, Non-ideal Circumstances’. A precursor paper was presented at the 2006 ALSP Annual Conference. Later versions were presented at the Philsoc Seminar, Australian National University and the 2007 Harvard Graduate Conference in Political Theory. I wish to thank the participants at these events for intense and stimulating discussions. I am also very grateful to Enrico Biale, Samantha Brennan, Alexander Brown, Ian Carter, Elizabeth Cripps, Cécile Laborde, Christian List, Saladin Meckled-Garcia, Thomas Pogge, Miriam Ronzoni, Zofia Stemplowska, Gabriel Wollner, the editor and the reviewers of *The Journal of Political Philosophy* for critical comments and constructive suggestions. Some of the contributions to the ECPR Workshop will appear in a special issue of *Social Theory and Practice* (forthcoming 2008).

¹ I am borrowing this broad definition of ideal theory from Sarah Williams Holtman, ‘Kant, ideal theory, and the justice of exclusionary zoning’, *Ethics*, 110 (1999), 32-58, at p. 34. Onora O’Neill offers extensive discussions of the notion of idealisation in *Towards Justice and Virtue* (Cambridge: Cambridge University Press, 1996), pp. 39-44, and ‘Abstraction, idealization and ideology in ethics’, *Moral Philosophy and Contemporary Problems*, ed. J.D.G. Evans (Cambridge: Cambridge University Press, 1988), pp. 55-69. John Rawls adopts a narrower notion of ideal theory, focusing on two idealisations in particular: full compliance and favourable conditions. In this article, I only discuss the notion of an ideal theory broadly construed, which includes, but is not limited to, Rawls’s narrower definition.

² John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1999 rev. ed.), pp. 4-8 (henceforth TJ); and *The Law of Peoples* (Cambridge, MA: Harvard University Press, 1999), pp. 4-5, 8-9 (henceforth LP).

³ Rawls, TJ, p. 125.

-
- ⁴ Ronald Dworkin, *Sovereign Virtue: The Theory and Practice of Equality* (Cambridge, MA: Harvard University Press, 2000), pp. 172 ff. (henceforth SV).
- ⁵ A recent formulation of this critique is offered by Colin Farrelly, 'Justice in ideal theory: a refutation', *Political Studies*, 55 (2007), 844-64. See also Amartya Sen's critique of what he calls 'transcendental' theories of justice in his 'What do we want from a theory of justice?' *Journal of Philosophy*, 103 (2006), 215-38.
- ⁶ G.A. Cohen, 'Facts and principles', *Philosophy and Public Affairs*, 31 (2003), 211-45. To be more precise, Cohen does not talk about fact-sensitive and fact-insensitive normative *theories*, but rather about fact-sensitive and fact-insensitive normative *principles* (of justice in particular). Since, however, principles are the main components of theories, Cohen's distinction naturally extends to different ways of theorising about justice.
- ⁷ My discussion here also benefits from Michael Blake's distinction between ideal institutional and non-institutional theorising in 'Distributive justice, state coercion, and autonomy', *Philosophy and Public Affairs*, 30 (2001), 257-96, esp. at pp. 263-4, n.7, and from Aaron James's distinction between practice-based and non-practice-based theorising, 'Constructing justice for existing practice: Rawls and the status quo', *Philosophy and Public Affairs*, 33 (2005), 281-316, at pp. 293-5. Cf. Andrea Sangiovanni's somewhat wider distinction between relational and nonrelational approaches to justice in 'Global justice, reciprocity, and the state', *Philosophy and Public Affairs*, 35 (2007), 3-39, at pp. 5-8. See also his 'Justice and the priority of politics to morality', *Journal of Political Philosophy*, 16 (2008), 137-64.
- ⁸ In offering this formulation in terms of power I am indebted to Aaron James's discussion in 'Power in social organization as the subject of justice', *Pacific Philosophical Quarterly*, 86 (2005), 25-49.
- ⁹ This approach is defended by G.A. Cohen in 'Facts and principles'. As Cohen himself remarks (p. 244, n.48), Richard Arneson seems to agree with certain aspects of his view. See R.J. Arneson, 'Luck egalitarianism and prioritarianism', *Ethics*, 110 (2000), 339-49, p. 345. Andrew Mason, 'Just

constraints’, *British Journal of Political Science*, 34 (2004), 251-68, and Adam Swift, ‘The value of philosophy in non-ideal circumstances’, *Social Theory and Practice* (forthcoming 2008) also approve of this form of theorising.

¹⁰ Colin Farrelly, for instance, proposes this continuum model in ‘Justice in ideal theory: a refutation’.

¹¹ This can be taken as a generalised description of what Cohen, ‘Facts and principles’, at p. 241, calls ‘principles of regulation’ which, he says, are the object of fact-sensitive theories. My understanding of the distinction between theories of type (1) and (2) has benefited from many conversations with Miriam Ronzoni, who explores it in depth in her doctoral thesis ‘Constructivism and the basic structure of society’ (Oxford D.Phil thesis, forthcoming 2008).

¹² G. A. Cohen, ‘Expensive taste rides again’, *Dworkin and His Critics*, ed. J. Burley (Oxford: Blackwell, 2004), pp. 3-29, at p. 18.

¹³ Cohen, ‘Facts and principles’, p. 243, emphasis added.

¹⁴ For a related discussion see Miriam Ronzoni and Laura Valentini, ‘On the meta-ethical status of constructivism: reflections on G.A. Cohen’s “Facts and principles”’, *Politics, Philosophy & Economics* (forthcoming).

¹⁵ So-called democratic or relational egalitarians can also be counted among the advocates of the fact-sensitive approach. See, e.g., Elizabeth Anderson, ‘What is the point of equality?’ *Ethics*, 109 (1999), 287-337, and Samuel Scheffler, ‘What is egalitarianism?’ *Philosophy and Public Affairs*, 31 (2003), 5-39.

¹⁶ John Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001), p. 41 (henceforth JFR).

¹⁷ John Rawls, *Political Liberalism* (New York: Columbia University Press, 1996), pp. 44 ff. (henceforth PL).

¹⁸ *Ibid.*, p. 46.

¹⁹ For an early statement of this conception of justification see Rawls, TJ, pp. 506 ff.

²⁰ Rawls, PL, p. 110. On the divergence between individuals’ priority judgements see TJ, pp. 39-40.

-
- ²¹ A similar description of Rawls's approach in terms of a 'public framework of thought' is offered by Sen in 'What do we want from a theory of justice?'
- ²² Dworkin, SV, p. 1. Notice that Dworkin uses the notion of political power much more narrowly than I do. Instead of seeing it as the power members of a political community exercise over one another through the operation of their institutions, he conceives of it in terms of the influence members have over political decisions. See *Ibid.*, p. 65 and ch. 4.
- ²³ *Ibid.*, p. 4, emphasis added.
- ²⁴ For a detailed taxonomy of ideal theory see Ingrid Robeyns, 'Ideal theory in theory and practice', *Social Theory and Practice* (forthcoming 2008). See also Charles W. Mills, "'Ideal theory" as ideology', *Hypatia*, 20 (2005), 165-84, esp. at pp. 166-7.
- ²⁵ In 'What do we want from a theory of justice?' Amartya Sen has challenged this claim, arguing that theories of justice can also be purely comparative. Such theories tell us how different societies ought to be ranked from the viewpoint of justice, without specifying what a perfectly just society would look like. I thank Christian Barry for drawing my attention to this issue.
- ²⁶ Zofia Stemplowska has recently argued in favour of a different technical definition of ideal theory, based on the particular function ideal theory is meant to perform. See her 'What's ideal about ideal theory?' *Social Theory and Practice* (forthcoming 2008).
- ²⁷ Onora O'Neill, for instance, argues that, unlike abstractions, idealisations can (and perhaps should) be avoided. For discussion see O'Neill, *Towards Justice and Virtue*, pp. 39-44. For scepticism about O'Neill's position see Miriam Ronzoni, 'Constructivisms: between abstraction and idealisation' (unpublished manuscript); Lisa H. Schwartzman, 'Abstraction, idealization, and oppression', *Metaphilosophy*, 37 (2006), 565-88, and Robert E. Goodin, 'Political ideals and political practice', *British Journal of Political Science*, 25 (1995), 37-56, esp. at pp. 40-45.
- ²⁸ Rawls, TJ, pp. 118-23.
- ²⁹ *Ibid.*, pp. 137-8.

³⁰ *Ibid.*, pp. 123-5.

³¹ Dworkin, SV, p. 172.

³² *Ibid.*, pp. 66 ff. For a helpful summary of Dworkin's theory to which I am indebted see Will Kymlicka, *Contemporary Political Philosophy: An Introduction* (Oxford: Oxford University Press, 2002), pp. 75-88.

³³ Dworkin, SV, pp. 67-8.

³⁴ *Ibid.*, pp. 73 ff.

³⁵ *Ibid.*, pp. 147-61 and 161-2 respectively.

³⁶ *Ibid.*, pp. 162-180.

³⁷ Of course, what counts as 'humanly achievable' is a controversial, largely speculative, matter.

³⁸ In so saying, I am not suggesting that a theory of justice cannot be 'incentive-compatible'. Indeed, one of the most discussed features of Rawls's difference principle is its attempted incentive-compatibility. Recall that the principle justifies inequalities so long as they are needed to benefit the worse-off, and these include inequalities generated to incentivise talented people to make occupational choices that maximise productivity. Compatibility of a theory of justice with a certain degree of human self-interestedness seems necessary given that justice, as a value, makes sense only because human beings are partly self-interested, as opposed to angelically altruistic. The point I am making in the main text is that, once the principles of justice have been defined (whatever their degree of incentive-compatibility) each individual in society is assumed to do her part in upholding them. For a critique of Rawls's incentives argument for inequality, see G.A. Cohen, *If You're an Egalitarian, How Come You're So Rich?* (Cambridge, MA: Harvard University Press, 2000), esp. chs 8 and 9. I am grateful to Christian List for bringing this issue to my attention.

³⁹ For discussion see Raymond Geuss, 'Neither history nor praxis', *Outside Ethics*, in R. Geuss (Princeton, NJ: Princeton University Press, 2005), pp. 29-39.

⁴⁰ For a helpful account of the ‘psychological, sociological, and epistemological’ constraints to be taken into account by a theory that aims to be practically effective see Joseph H. Carens, ‘Realistic and idealistic approaches to the ethics of migration’, *International Migration Review*, 30 (1996), 156-70. On related issues, see also Juha Räikkä, ‘The feasibility condition in political theory’, *The Journal of Political Philosophy*, 6 (1998), 27-40; Sanjay Reddy, ‘The role of apparent constraints in normative reasoning: a methodological statement and application to global justice’, *The Journal of Ethics*, 9 (2005), 119-25, and Robert E. Goodin’s discussion of what he calls the ‘accessibility critique’ of ideals in ‘Political ideals and political practice’, pp. 39-40.

⁴¹ Immanuel Kant, ‘On the common saying: this may be true in theory, but it does not apply in practice’, *Kant’s Political Writings*, ed. H.S. Reiss (Cambridge: Cambridge University Press, 1970), pp. 61-92, at p. 61.

⁴² This point is very well made by O’Neill in ‘Abstraction, idealization and ideology in ethics’, at p. 59.

⁴³ See, e.g., the complex argument through which Sarah Williams Holtman shows how Kantian ideal theory might offer guidance in establishing the legitimacy of real-world policies in ‘Kant, ideal theory, and the justice of exclusionary zoning’.

⁴⁴ Kant, ‘On the common saying: this may be true in theory, but it does not apply in practice’, p. 61.

⁴⁵ At points, Annette Baier does seem to advance such a radical critique to normative theorising in ‘Theory and reflective practices’, *Postures of The Mind: Essays on Mind and Morals*, in A. Baier (London: Methuen, 1985), pp. 207-27.

⁴⁶ *Ibid.*, p. 225.

⁴⁷ For a sustained discussion of this difficulty see Michael Phillips, ‘Reflections on the transition from ideal to non-ideal theory’, *Noûs*, 19 (1985), 551-70.

⁴⁸ For similar examples see *Ibid.*, pp. 559-60.

⁴⁹ O’Neill, *Towards Justice and Virtue*, p. 41. Notice, however, that O’Neill’s overall critique of idealisation does not so much focus on the issue of guidance, but rather on the difficulty of justifying idealised

premises without appealing to controversial metaphysical claims. For O'Neill's remarks on issues of guidance see 'Abstraction, idealization and ideology in ethics'.

⁵⁰ Baier, 'Theory and reflective practices', p. 210, emphasis added.

⁵¹ I point this out because critics often fail to distinguish between GC2 and GC3.

⁵² Thomas McCarthy, 'Political philosophy and racial injustice: from normative to critical theory', *Pragmatism, Critique, Judgement*, eds. S. Benhabib and N. Fraser (Cambridge MA: MIT Press, 2004), pp. 147-70, at p. 165, emphasis original. For a similar claim on the difficulty in bridging the gap between ideal and nonideal theory see Phillips, 'Reflections on the transition from ideal to non-ideal theory', p. 551. Joel Feinberg also offers some sceptical remarks on the guidance capacity of Rawls's ideal theory, particularly with reference to the case of civil disobedience, in 'Duty and obligation in the non-ideal world', *Journal of Philosophy*, 70 (1973), 263-75.

⁵³ Mills, "'Ideal theory'" as ideology'.

⁵⁴ My argument is thus limited in scope. I only discuss deontological liberal approaches to justice and do not consider whether the 'guidance critique' would also apply to other types of outlooks. For instance, as David Grewal and Gabriel Wollner have (independently) pointed out to me, consequentialist theories of justice might be immune to the challenges discussed here.

⁵⁵ Rawls, JFR, p. 66.

⁵⁶ For a general discussion of racial and sexual discrimination containing helpful methodological reflections on the relation between 'ideals' and 'social realities' see Richard A. Wasserstrom, 'Racism, sexism and preferential treatment: an approach to the topics', *Contemporary Political Philosophy: An Anthology*, eds. R. Goodin and P. Pettit (Oxford: Blackwell, 1997), pp. 579-604.

⁵⁷ Susan Moller Okin, *Justice, Gender, and the Family* (New York: Basic Books, 1989), esp. pp. 89-109.

⁵⁸ Rawls, JFR, pp. 162-8.

⁵⁹ *Ibid.*, p. 167.

-
- ⁶⁰ This point is defended and discussed in Thomas W. Pogge, ‘On the site of distributive justice: reflections on Cohen and Murphy’, *Philosophy and Public Affairs*, 29 (2000), 137-69, esp. at p. 165. The later Rawls (JFR, p. 12) is explicit in wanting the notion of the basic structure to be context-sensitive. He states ‘[w]ere we to lay down a definition of the basic structure that draws sharp boundaries ... we would also risk wrongly prejudging what more specific or future conditions may call for, thus making justice as fairness unable to adjust to different social circumstances’.
- ⁶¹ Rawls seems to be more radical on this point, suggesting that how to best realise women’s equality ‘in particular historical conditions is not for political philosophy to decide’, Rawls, JFR, p. 167. Here I am making the weaker claim that this is not for political philosophy *alone* to decide.
- ⁶² Rawls, TJ, p. 327.
- ⁶³ *Ibid.* In particular, Rawls thinks that violations of the difference principle are the most difficult injustices to detect, while violations of the first principle and ‘blatant violations’ of fair equality of opportunity are more likely to be ‘obvious to all’, *Ibid.*, pp. 326-7.
- ⁶⁴ I am grateful to an anonymous referee for raising this concern.
- ⁶⁵ I wish to thank Miriam Ronzoni for bringing this point to my attention. Ronzoni offers a sustained defence of Rawls’s basic structural approach in ‘Constructivism and the basic structure of society’.
- ⁶⁶ Roland Pierik and Ingrid Robeyns, ‘Resources versus capabilities: social endowments in egalitarian theory’, *Political Studies*, 55 (2007), 133-52, at p. 141, emphasis original. For a critique of Dworkin’s methodology see also Philip Abbott, ‘Review of *Sovereign Virtue*’, *Rhetoric & Public Affairs*, 4 (2001), 576-8.
- ⁶⁷ Pierik and Robeyns, ‘Resources versus capabilities’, p. 142.
- ⁶⁸ Dworkin, SV, p. 165.
- ⁶⁹ In spite of her critique of Dworkin, Robeyns (‘Ideal theory in theory and practice’) now recognizes the legitimacy of such idealisations for the purpose of designing an ideal theory of justice.

⁷⁰ Critics sometimes recognize this point, but lament that, apart from sporadic remarks (for instance, Dworkin says ‘If proper civil rights laws were enacted and enforced, the influence of prejudice ... would wane’, SV, pp. 345-6), Dworkin seldom discusses the question of how to bring socially generated inequalities to an end *directly* in connection with his ideal theory. See Pierik and Robeyns, ‘Resources versus capabilities’, pp. 143-6 and Robeyns, ‘Ideal theory in theory and practice’. This may very well be true but, as I already noticed, the absence of explicit discussion does not mean that a fruitful discussion could not be developed on the basis of Dworkin’s theory.

⁷¹ Dworkin, SV, chs 11 and 12.

⁷² For discussion see Thomas Nagel, ‘Equal treatment and compensatory discrimination’, *Equality and Preferential Treatment*, ed. M. Cohen, T. Nagel and T. Scanlon (Princeton, NJ: Princeton University Press, 1977), pp. 3-18.

⁷³ Though this accusation is probably unfair in the case of Dworkin, who devotes about half of *Sovereign Virtue* to the ‘practice’ of equality.

⁷⁴ Rawls, LP, pp. 3, 10, emphasis original.

⁷⁵ In Rawls’s words (*Ibid.*, p. 6, n.7), as a realistic utopia the Law of Peoples ‘sets limits to the reasonable exercise of power’.

⁷⁶ For Rawls’s list see *Ibid.*, p. 37.

⁷⁷ *Ibid.*, p. 4. More precisely, burdened societies are societies which ‘lack the political and cultural traditions, the human capital and know-how, and, often, the material and technological resources needed to be well-ordered’. *Ibid.*, p. 106.

⁷⁸ Allen Buchanan, ‘Rawls’s Law of Peoples: rules for a vanished Westphalian world’, *Ethics* 110 (2000), 697-721, esp. at pp. 716-20.

⁷⁹ Simon Caney, ‘Cosmopolitanism and the Law of Peoples’, *Journal of Political Philosophy*, 10 (2002), 95-123, at pp. 118-9.

⁸⁰ Rawls, LP, p. 89.

-
- ⁸¹ Thomas W. Pogge, 'Moral universalism and global economic justice', *Politics, Philosophy & Economics*, 1 (2002), 29-58, esp. at pp. 46-9, reprinted in Pogge, *World Poverty and Human Rights* (Cambridge: Polity, 2002), ch. 4. For an overall critical appraisal of *The Law of Peoples* see also Pogge's 'Critical study: Rawls on international justice', *Philosophical Quarterly*, 51 (2001), 246-53.
- ⁸² On this see, e.g., David Singh Grewal, 'Network power and globalization', *Ethics and International Affairs*, 17 (2003), 89-98; Darrell Moellendorf, 'The World Trade Organization and egalitarian justice', *Metaphilosophy*, 36 (2005), 145-62; and Richard H. Steinberg, 'In the shadow of law or power? Consensus-based bargaining and outcomes in the GATT/WTO', *International Organizations*, 56 (2002), 339-74.
- ⁸³ Thomas W. Pogge, "'Assisting" the global poor', *The Ethics of Assistance: Morality and the Distant Needy*, ed. D.K. Chatterjee (Cambridge: Cambridge University Press, 2004), pp. 260-88.
- ⁸⁴ Notice that Rawls is not unaware that existing liberal states are far less well-ordered than his 'peoples' (LP, pp. 48, 53), yet he seems to ignore this in his discussion of non-ideal theory.
- ⁸⁵ Will Kymlicka, *Multicultural Citizenship: A Liberal Theory of Minority Rights* (Oxford: Oxford University Press, 1995) (hereafter MC).
- ⁸⁶ Will Kymlicka, 'Do we need a liberal theory of minority rights? Reply to Carens, Young, Parekh and Forst', *Constellations*, 4 (1997), 72-87, at p. 75.
- ⁸⁷ Kymlicka, MC, pp. 214, n.11, 109.
- ⁸⁸ *Ibid.*, p. 152, emphasizes original.
- ⁸⁹ *Ibid.*, ch. 8.
- ⁹⁰ *Ibid.*, pp. 165, 169.
- ⁹¹ See, e.g., Brian Barry, *Culture and Equality* (Cambridge: Polity Press, 2001), ch. 4. See also Chandran Kukathas, 'Multiculturalism as fairness: Will Kymlicka's *Multicultural Citizenship*', *Journal of Political Philosophy*, 5 (1997), 406-27, at p. 420, and Geoffrey Brahm Levey, 'Equality, autonomy, and cultural rights', *Political Theory*, 25 (1997), 215-48, esp. at pp. 227 ff.

⁹² For a sustained critical discussion of the limitations of Kymlicka's ideal theoretic approach to which I am indebted see Adrian Favell, 'Applied political philosophy at the Rubicon: Will Kymlicka's multicultural citizenship', *Ethical Theory and Moral Practice*, 1 (1998), 255-78.

⁹³ Kymlicka, MC, p. 165.

⁹⁴ As Cécile Laborde has suggested to me, this clearly emerges from the fact that Kymlicka defends national minorities because they provide a context of choice. This implicitly assumes minorities to be *already* liberal (just). Otherwise, how could they be a context of *choice*?

⁹⁵ Kymlicka, MC, p. 164.

⁹⁶ As Zofia Stemplowska has helpfully pointed out to me, right-libertarians would disagree with this claim, arguing that egalitarians like Rawls and Dworkin have an implausibly idealised conception of the state. For right-libertarians the complex state apparatus necessary to realize egalitarian policies would inevitably fall prey to corruption and power abuses, thus hindering, rather than promoting, the pursuit of justice. My argument in this article is entirely internal to liberal egalitarianism and thus does not respond to this libertarian critique.

⁹⁷ Rawls, LP, p. 42, n.52.

⁹⁸ Simon Caney, 'Cosmopolitanism and the Law of Peoples', p. 118, and Thomas W. Pogge, 'The incoherence between Rawls's theories of justice', *Fordham Law Review*, 72 (2004), 1739-59, at p. 1751. Pogge offers a detailed critical analysis of the structural differences between Rawls's domestic and international theories of justice, which has informed my own discussion here. In particular, he argues that while Rawls's domestic theory is three-tiered and institutional, his international theory is two-tiered and interactional. In Pogge's view, these structural asymmetries explain much of what is problematic about *The Law of Peoples*. For further discussion of Pogge's view in the relation to the argument presented here, see footnote 106 of this article.

⁹⁹ Interestingly, some scholars have emphasised this aspect of Rawls's theory to defend it against cosmopolitan objections. In their view, since Rawls's principles are meant to apply in 'ideal theory'

– where no power abuses occur – they should not be objected to on the grounds that they are misleading (or unhelpful) in real-world circumstances. Obviously, this defence of *The Law of Peoples* only succeeds if we reject the claim that any sound normative theory should be action-guiding. For this line of argument see, e.g., Stephen Macedo, ‘What self-governing peoples owe to one another: universalism, diversity, and *The Law of Peoples*’, *Fordham Law Review*, 72 (2004), 1721-38, esp. at p. 1727 and Samuel Freeman, ‘The Law of Peoples, social cooperation, human rights, and distributive justice’, *Social Philosophy and Policy*, 23 (2006), 29-68, at pp. 32-3.

¹⁰⁰ I borrow this expression from Phillips, ‘Reflections on the transition from ideal to non-ideal theory’.

¹⁰¹ Buchanan, ‘Rawls’s Law of Peoples: rules for a vanished Westphalian world’.

¹⁰² Interestingly, this is acknowledged even by advocates of non-ideal theory, for instance by Mills, ‘“Ideal theory” as ideology’, at p. 171, when he says that ‘nonideal theory can and does appeal to an ideal’.

¹⁰³ In connection with this point, Ingrid Robeyns and Miriam Ronzoni have suggested to me that a dimension along which we may distinguish between good and bad idealisations is their flexibility, that is the extent to which they constitute an integral part of a theory. The more idealisations are integral to a theory, the harder it is for the theory to adjust to its ‘real-world’ object and to provide valuable tools for studying it. Robeyns further discusses how to distinguish between good and bad idealisations in ‘Ideal theory in theory and practice’.

¹⁰⁴ I owe this point to a very helpful discussion with Christian List.

¹⁰⁵ O’Neill, *Towards Justice and Virtue*, p. 41.

¹⁰⁶ Pogge’s distinction between two-tiered and three-tiered theories (introduced at footnote 98 of this article) helps explain this aspect of Rawls’s approach to international justice. In particular, Pogge claims that while Rawls’s domestic theory of justice is three-tiered – the parties in the original position (1) select principles (2) to govern the basic structure of society (3) – the ‘Law of Peoples’ is two-tiered: the representatives of peoples (1) select principles for guiding peoples’ conduct (2). In

light of our discussion, we might say that while three-tiered theories ensure a separation between the justification of principles (1 and 2) and their application to the subject of justice (3), this does not happen with two-tiered ones, such as the Law of Peoples. This is why idealisations that would be appropriate at the level of its justification end up characterizing the subjects to which Rawls's theory of international justice is meant to apply. See Pogge, 'The incoherence between Rawls's theories of justice'. On related themes see Lea L. Ypi, 'Getting it backwards: ideal and non-ideal in recent debates on global justice' (unpublished manuscript).

¹⁰⁷ O'Neill also briefly remarks on the use of idealisations in theoretical, as opposed to practical, reasoning, in *Towards Justice and Virtue*, pp. 41-2.

¹⁰⁸ I am grateful to Robert E. Goodin for suggesting these examples.

¹⁰⁹ See Daniel M. Hausman, 'Philosophy of economics', *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2006 edn); available at: <http://plato.stanford.edu/archives/sum2006/entries/economics/>.

¹¹⁰ In recent years, however, behavioural economists have considered more empirically informed behavioural assumptions and investigated how these affect standard results in economic theory.