On the Apparent Paradox of Rawlsian Ideal Theory*

Laura Valentini (UCL)
l.valentini@ucl.ac.uk

This is just a draft and comments are most welcome but do not quote, cite or circulate without permission

Introduction

It is often remarked that contemporary liberal egalitarian conceptions of justice are developed in ‘ideal theory’. In spite of its popularity, the notion of an ideal theory tends to be employed somewhat loosely, to indicate any theory constructed at a high level of abstraction, under assumptions which render existing social reality significantly “simpler and better”1 than it actually is. Admittedly general, this definition is well suited to 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 ideal world.4 That is, it is designed assuming 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 widely spread, this idealising trend has increasingly 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 likely to be ill-suited to guide action in the real world, where such conditions no longer hold. Given that a capacity for guidance is widely regarded as a necessary attribute of any plausible normative

---

1 I wish to thank Cécile Laborde, Enrico Biale and Elizabeth Cripps for reading and helpfully commenting on an earlier draft of this paper, and Miriam Ronzoni for instructive and inspiring discussions on many of the issues investigated here. The argument in section 5 draws on a paper I presented at the 2006 ALSP Annual Conference, and benefits from the comments and suggestions I received on that occasion.

2 I am borrowing this broad definition of ideal theory in terms of simplification and improvement from Sarah Williams Holtman’s “Kant, Ideal Theory and the Justice of Exclusionary Zoning”, *Ethics*, 110 (1) (1999), 32-58, 34. John Rawls famously adopts a narrower notion of ideal theory, which I shall discuss in section 2. For now, I am only concerned with ideal theory intuitively understood and loosely construed.

3 TJ, 125.

theory, the ‘guidance critique’ seems to pose a serious threat to liberal egalitarian thinking as a whole.\textsuperscript{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: for how could we even formulate judgements about the justice and injustice of society, let alone promote institutional reform, without having at least a sketchy, yet coherent, \textit{ideal} of what a just society would look like? This being the case, liberal egalitarians find themselves in the contradictory situation where their ideal theories appear to be both \textit{unable} to guide action and \textit{indispensable} to guide action. This, which I call ‘the paradox of ideal theory’, can be articulated as follows:

\begin{itemize}
\item[a)] (Liberal egalitarian) theories of justice aim to be action-guiding.
\item[b)] Theories of justice are necessarily ideal.
\item[c)] Ideal theories fail to guide action in non-ideal circumstances.
\end{itemize}

Is there a way out of this paradox? In this paper I argue that there is. I reach this conclusion through a two-stage argument. In the first I articulate and examine claims a), b) and c), with a view to clarifying the terms of the discussion and putting the target of the present investigation into sharper focus. In fact, I have deliberately formulated the paradox very loosely, because what renders the question of ideal theory particularly intractable is, at least in part, a matter of conceptual vagueness and confusion. The first task of this paper is thus to offer a first step towards clarification, separating what might be regarded as genuine questions about the validity of ideal theorising, from questions which, on scrutiny, turn out to have a different object. More specifically, in section 1 I distinguish between two liberal egalitarian approaches to justice: institutional and non-institutional. I then restrict the scope of my discussion to the former, particularly focusing on John Rawls as its most eminent proponent. In section 2, I show that claim b), which affirms the unavoidability of ideal theorising, is warranted and in section 3 I spell out the gap critique in greater detail, distinguishing between theoretically challenging and unchallenging versions of it.

Having narrowed down the scope of my investigation and further qualified the terms of the paradox, I turn to the second part of the paper. In sections 4 and 5 I assess the validity of the guidance critique in relation to two ‘case studies’, drawn, respectively, from John Rawls’s domestic and international theories of justice. I argue that, whilst the critique does not apply to the former, it has bite on the latter. This suggests that there is nothing wrong with ideal theory \textit{per se}, but rather that there can

\textsuperscript{5} A recent formulation of this critique is offered by Colin Farrelly in his “Justice in Ideal Theory: A Refutation”, forthcoming in \textit{Political Studies}. 

ECPR Joint Sessions, 7-12 May 2007
be good and bad forms of ideal theorising: whether certain idealisations are unwarranted, hence whether a normative theory is vulnerable to the guidance critique, is something to be judged on a case by case basis, in the light of the specific question the theory itself addresses.

1. Claim a): Justice and Guidance

As I said at the outset, 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 affects liberal egalitarianism across the board: prominent theorists such as John Rawls, Ronald Dworkin, G.A. Cohen and Richard Arneson, all qualify as appropriate targets. However, this would be too hasty a conclusion. Whether a specific liberal egalitarian outlook has to confront the paradox of ideal theory depends on its endorsement of claim a), namely the claim that conceptions of justice aim to be action-guiding. At first, such a claim appears unquestionable: how can a normative theory lack such an aspiration? Contrary to expectations, however, some forms of liberal egalitarianism are entirely 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, precisely indicating what sort of theory of justice may be troubled by the paradox.

To do so, I begin by distinguishing, with Michael Blake, between two types of ideal theorising, both of which, I argue, are present in contemporary liberal egalitarian thinking: institutional and noninstitutional.6 The main difference between these approaches lies in the attitude they take towards existing institutions and practices. Ideal institutional theories, says Blake, ask “what the institutions we currently have would have to do to be justified”, under favourable conditions and assuming all participants did their fair share.7 Typically, theories of justice focus on the justification of coercive institutions and practices, 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. They work bottom-up, for they hold that the scope and content of obligations of justice cannot be determined prescinding from an account of the type of power-relations we are actually involved in, through participating in common institutions.8 For institutional theorists, one might say, the question of justice only arises when such power-relations are in play.9

9 In offering this formulation in terms of power I am indebted to Aaron James’s “Power in Social Organization as the Subject of Justice”, Pacific Philosophical Quarterly, 86 (1) (2005), 25-49.
Ideal noninstitutional theories, on the other hand, rest on much thinner empirical assumptions (if any) and, abstracting away from existing practices and institutions, attempt to define the value of justice ‘in its pure form’, to then ask “what institutions we would endorse if we were starting from scratch.”

Typically, they think of justice as a fundamental moral value, expressed by some principle of distributive equality, the nature of which is insensitive to existing facts, including the ‘fact’ of political power exercised within (and through) social practices and institutions.

Even though institutional and non-institutional approaches may be both described as ‘ideal’ in that they both abstract away from some (or all) existing facts, they clearly presuppose altogether different ethical outlooks and conceive of the question of justice in completely different terms. Schematically put:

a. Non-Institutional theories ask: 1) What is the value of justice? 2) What would an ideal social order, one that optimally realises justice – in conjunction with other values – look like?

b. Institutional theories ask: What principles should govern the exercise of political power, given the forms in which such a power is currently being exercised?

Non-institutional theorists tend to focus on question 1 and, at least to my knowledge, they hardly ever go so far as to consider how their theories would generate duties and obligations for existing agents (question 2). In fact, non-institutional theorists do not seem to be at all troubled by the issue of guidance. As G.A. Cohen, perhaps the most radical proponent of a non-institutional approach asserts, justice might “not [be] something that the state, or, indeed, any other agent, is in a position to deliver.” Rather than telling us “what we should do”, principles of justice express “what we should think even when what we should think makes no practical difference.” This being the case, there is reason to leave non-institutional theorists – at least those who share Cohen’s view – 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 placed on the validity of a conception of

---

12 Aaron James suggests that this type of ideal theory contains an “optimality condition”. See his “Constructing Justice for Existing Practice: Rawls and the Status Quo”, 294-295.
justice, and are therefore left unmoved by the paradox.\textsuperscript{15} This only causes trouble to liberal egalitarian institutional theorists, of whom John Rawls is undoubtedly the most illustrious and criticized exponent.\textsuperscript{16}

His approach to justice is fully in line with the institutional paradigm as I have described it.\textsuperscript{17} 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. His theory articulates those principles any such society must follow in distributing benefits and burdens, if its coercive power is to be justified in the eyes of everyone. The idea is that citizens, conceived of as free and equal, should be given reasons to abide by the rules through which they mutually coerce one another, beyond the mere fact that they cannot avoid being so coerced because societal membership is not voluntary.\textsuperscript{18} As Rawls puts it, the question ‘justice as fairness’ attempts to answer is: “if the fact of reasonable pluralism always characterizes democratic societies and if political power is indeed the power of free and equal citizens, in the light of what reasons and values – of what kind of conception of justice – can citizens legitimately exercise that coercive power over one another?”\textsuperscript{19} His principles of justice, then, express those reasons and values which should be appealed to in order to determine the morally appropriate distribution of scarce resources within society, which in turn defines the structure of citizens’ power-positions and power-relations. When such principles are violated, instead of a society of equals who \textit{legitimately coerce} one another through participating in common institutions, we have a political system where some are \textit{oppressed} by others.

Second, Rawls’s conception of political philosophy is inherently concerned with the practical. As he says, the idea of “‘reasonable justification” is to be understood “as a practical and not as an epistemological or a metaphysical problem.”\textsuperscript{20} He sees the need for engaging in highly abstract (that is to say ideal) theorising as resting on the breakdown of “our shared political understandings”.\textsuperscript{21} It is only

\textsuperscript{15} Of course, this observation raises the deeper meta-ethical question of whether it makes sense to understand the nature of conceptions of justice in this way. However interesting, this question clearly differs from the one I have set out to tackle in this paper, to which I now turn.\textsuperscript{16} Also so-called democratic or relational egalitarians, such as Samuel Scheffler and Elizabeth Anderson, can be taken to follow the institutional paradigm, for they conceive of a just society as a society freed from oppression – i.e. one where political power is legitimately exercised – and where citizens relate to one another as equals. See Elizabeth Anderson, “What is the Point of Equality?”, \textit{Ethics} 109 (2) (1999), 287-337; and Samuel Scheffler, “What is Egalitarianism?”, \textit{Philosophy and Public Affairs} 31 (1) (2003), 5-39. Notice that there may be other forms of liberal egalitarianism which, in spite of their not being ‘institutional’ in the way I indicated, aspire to be action-guiding. This twofold taxonomy is not meant to be exhaustive, but only to articulate an important distinction, one that captures a sufficiently wide number of liberal egalitarian theorists for it to be of general interest.\textsuperscript{17} I am here considering Rawls’s theorising as a coherent whole. Interpretive disputes as to the continuity of his views from \textit{A Theory of Justice} to later writings do not bear on the present discussion: an institutional approach, a focus on ideal theorising and a conviction that political philosophy should be action-guiding have marked Rawls’s thinking all along.\textsuperscript{18} Interpretations of Rawls placing emphasis on this point are offered by Blake, “Distributive Justice, State Coercion, and Autonomy”, 283; A.J. Julius, “Basic Structure and The Value of Equality”, \textit{Philosophy and Public Affairs}, 31 (4) (2003), 321-355.\textsuperscript{19} John Rawls, \textit{Justice as Fairness, A Restatement} (Cambridge MA, Belknap Press Harvard, 2001), 40-41 henceforth JFR.\textsuperscript{20} John Rawls, \textit{Political Liberalism}, (New York, Columbia University Press, 1996), 44, henceforth PL.\textsuperscript{21} PL, 44.
when criteria for justification become actually contested within society that philosophical reflection at a high level of abstraction is called for. For instance, there is no need to engage in ideal theorising to establish whether torture is wrong. Ideal theorising is not meant to help us with such a clear-cut case. As Rawls says, “The work of abstraction … is not gratuitous: not abstraction for abstraction’s sake. Rather, it is a way of continuing public discussion when shared understandings of lesser generality have broken down.” The task of political philosophy and ideal theory is to reduce (hopefully resolve) such disagreements, offering a “public framework of thought” from within which judgements about the justice of the current distribution of resources within society may be carried out. In Rawls’s words “[e]xisting institutions are to be judged in the light of this [i.e. the ideal] conception and held to be unjust to the extent that they depart from it without sufficient reason.” Ideal theory provides a tool for understanding and appraising reality from the moral viewpoint of justice. “[The reason for beginning with ideal theory”, says Rawls, “is that it provides … the only basis for the systematic grasp of [the] more pressing problems” arising in the non-ideal circumstances of actual political and social life. In short, ideal theory must come first because it enables us to make sense of what it is for political power to be justified in practice.

This concludes my analysis of claim a). As we have seen, not all liberal egalitarian theorists endorse the claim in question, hence not all of them are troubled by the paradox. Those who endorse such a claim, most famously John Rawls, think of principles of justice as offering criteria for the justification of the exercise of political power, and believe a theory of justice should be action-guiding in the sense of being capable of offering a common framework of thought from within which to assess, criticize and reform the way power is exercised within society. In Aaron James’s words, Rawls is “constructing justice for existing practice”. With these ideas clearer in mind, we can now move on to claim b, and give a closer look at the notion of an ideal theory, which I have so far left admittedly vague.

2. Claim b): Are Theories of Justice Necessarily Ideal?

The second component of our paradox states that ‘Theories of justice are necessarily ideal.’ To evaluate the tenability of this claim we first need a more precise understanding of what it is for a theory of justice to be ideal, and then to consider whether theories of justice must necessarily be ideal in that way. Those who are familiar with Rawls’s work are likely to point out that he holds a much narrower

---

22 PL, 45-46.
23 For an early statement of this conception of justification see TJ, 506ff.
24 PL, 110. See also TJ, 39-40, on the divergence between individuals’ priority judgements.
25 TJ, 216.
26 TJ, 8.
27 An interesting discussion of the function of Rawls’s ideal theory is offered by Jon Mandle in his “Justice, Desert and Ideal Theory”, Social Theory and Practice, 23 (3) (1997), 399-245, esp. 412-421.
28 James, “Constructing Justice for Existing Practice: Rawls and the Status Quo”.

6
understanding of the notion of an ideal theory than the one I have adopted so far. Whilst I have defined ideal theory as theory that significantly abstracts from existing facts, thereby improving and simplifying social reality, Rawls defines it as theory developed assuming full-compliance and reasonably favourable conditions. But are these two definitions of ideal theory all that different? I believe not. In fact, Rawls’s seemingly distinctive notion of ideal theory is not only compatible with the wider definition I have given at the outset, but it also offers a conceptually compelling understanding of what an ideal theory of justice must be. To illustrate this point, I shall take up and discuss the stipulations of favourable conditions and of full-compliance in turn.

Given our purposes, the assumption of reasonably favourable conditions can be dealt with relatively easily, for it simply amounts to the presupposition that, thanks to its political culture, level of economic development and resource endowment, society is capable of satisfying its citizens’ basic socioeconomic needs, thereby guaranteeing the “effective establishment” of their basic liberties. Since Rawls’s intended indicanda, i.e. liberal democratic societies such as the United States, may plausibly be regarded as capable of fulfilling their citizens’ basic needs, the stipulation of favourable conditions does not qualify as an idealisation strictly understood: it does not make things better, or simpler, than they actually are. What is really distinctive about Rawls’s notion of ideal theory is the stipulation of full compliance.

A theory elaborated under such a stipulation designs principles of justice presupposing that all agents falling under their purview will comply with them – the agent being the basic structure of a democratic society in Rawls’s case. As a result, the theory sketches the features of a fully just, hence ideal, social order. This is the kind of society we would obtain if we were to remove all injustices affecting existing democracies, where his principles are certainly not fully complied with.

Notice however, that, so understood, the Rawlsian notion of an ideal theory is easily traceable to the idea of selective abstraction from existing facts. That is, Rawls’s ideal theory is selectively fact-insensitive in that, assuming full-compliance, it attempts to offer a portrait of an injustice-free social order, one where no arbitrary discriminations in the adjudication of citizens’ competing interests occur. In fact, what else could an ideal theory of justice be, if not a theory that attempts to describe an injustice-free social order, namely a social order where all agents comply with principles of justice?

Far from being idiosyncratic, then, Rawls’s understanding of ideal theory offers a perfectly general and inescapable account of this notion, one that proves not only consistent with, but further qualifies, the general definition in terms of fact-insensitivity offered at the outset. With the notion of an ideal

29 TJ, 8, 215ff.; LP, 5.
theory clearer in mind, it remains for us to test whether conceptions of justice must be elaborated in ideal theory. The argument I have just offered suggests that this is indeed a conceptual truth: by definition, a theory that tells us what is just is one that abstracts away from all existing injustices and is therefore ideal. Is this explanation sufficient to establish the inescapability of ideal theorising? Some might say it is not. They might object that, in saying that ideal theorising abstracts from all existing injustices, I have mistaken the operation of abstraction for that of idealisation. The target of critics of ideal theory, the objection would continue, is not the former, but the latter theoretical move: ideal theory fails not because it abstracts but because it idealises.

One of the most articulate advocates of idealisation-free, yet abstract, theorising is Onora O’Neill.\(^{31}\) She distinguishes between abstraction and idealisation and argues that, whilst the former “is a matter of bracketing, but not of denying, predicates that are true of the matter under discussion”,\(^ {32}\) the latter “ascribes predicates that are false of the case at hand, and so denies predicates that are true of that case”\(^ {33}\). Typically, such predicates also have a positive axiological connotation, describing ideals the theorist believes the world should conform to.\(^ {34}\) Whilst abstraction, O’Neill claims, is an unavoidable and theoretically sound mechanism, idealisation is a potentially misleading move and should therefore be avoided. But can it be avoided? If it were possible to abstract without thereby also idealising, as O’Neill suggests, then claim b) would be false and the paradox easily overcome: not all normative theorising would need to be ideal. This would be a comforting conclusion but, unfortunately, it does not hold: as O’Neill defines it, idealisation is indeed inescapable in normative theorising.

As critics have pointed out,\(^ {35}\) the distinction between abstraction and idealisation in normative theorising turns out to rest on shaky grounds. For instance, if we go back to Rawls’s theory, we can describe it using the language of abstraction and that of idealisation interchangeably.\(^ {36}\) As argued by Miriam Ronzoni, when constructing the original position, we might say either that Rawls assumes that citizens are equal (which is a false, ‘idealised’ assumption) or that he brackets off (i.e. abstracts from) differences that are claimed to be morally arbitrary, such as one’s social class, natural endowments and personal preferences, and should therefore have no bearing on the distribution of benefits and burdens within society. Once such facts are bracketed off, their existence is denied and reasoning proceeds as if no individuals in society were subject to differentiation according to sex, class, natural endowments etc.,

---


\(^{32}\) O’Neill, Towards Justice and Virtue, 40, emphasis in original.

\(^{33}\) O’Neill, Towards Justice and Virtue, 41.

\(^{34}\) See O’Neill, “Abstraction, Idealisation and Ideology in Ethics”, 56.


\(^{36}\) In making this argument I am indebted to Ronzoni’s critical appraisal of O’Neill’s distinction, and how this applies to Rawls’s construal of the original position in her “Constructivisms: Between Abstraction and Idealisation”.

8
when in fact, they are. Similar considerations hold for other assumptions, such as the parties’ full rational capacities and physical abilities (which could be re-described as a matter of bracketing off individuals’ occasional flaws in reasoning, and physical disabilities). By assuming fairness and equality, which are ‘false’ of the subject matter at hand – specifically of actual democratic societies – the theorist brackets off arbitrariness and unfairness.

At a more general level, as recently pointed out by Lisa Schwartzman, abstraction, i.e. the operation of bracketing off, always involves some “selective omission”, that is, taking decisions about what should be kept for the sake of elaborating a theory of justice and what should be excluded.37 So long as bracketing off involves selecting (i.e. deciding what matters for the purposes of theorising), some idealisation (i.e. departure from reality by denying some of its actual features) is unavoidable.38 One selection of facts can certainly be better than another given the aim of the theory, but it is not possible to avoid selecting altogether.

In particular, the selective omissions required by any plausible theory of justice will always yield some ideal to which we should aim. A society where benefits and burdens are distributed in a way that is insensitive to morally arbitrary facts – i.e. a society which justly distributes social goods – represents an ideal actual societies should try to conform to. What we obtain when theorising about justice are, to put it in Robert Goodin’s words “both ideals and abstractions, at one and the same time”39.

We have thus established that claim b) is warranted. Theories of justice are necessarily ideal: by definition they attempt to offer a set of principles which, if fully complied with, would generate an injustice-free society, a worthy ideal towards which we should aspire. It is now time to turn to the very heart of this paper and explore more in detail the substance of the guidance critique. In keeping with the Rawlsian focus announced at the outset, I consider the guidance critique as this applies to his theory of justice.

3. Claim c): The Guidance Critique and Its Interpretations

As I said at the outset, ideal theorising has been repeatedly and increasingly 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 shall argue, are not theoretically problematic, whilst the third is.

GC1: It might be argued that ideal theories lack a capacity for guidance because they fail to motivate agents in the real world: in practice, they do not ‘work’. For instance, as Raymond Geuss notes, “as

37 Schwartzman, “Abstraction, Idealization, and Oppression”, 570.
38 A similar point is also made by Ronzoni in “Constructivisms: Between Abstraction and Idealisation”. See also Schwartzman, “Abstraction, Idealization, and Oppression”, 573.
39 Goodin, “Political Ideals and Political Practice”, 41.
Rawls’s purportedly egalitarian theory became more entrenched and more highly elaborated, social inequalities in fact increased drastically in virtually all industrialized countries”.

Though sadly true, this observation misses its mark. The fact that people lack the political will to act in accordance with principles of justice is certainly to be regretted, yet it hardly counts as an objection to the theoretical 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. If we were to judge its validity ex post, by reference to its actual impact, then the best theory would be a rationalization of the status quo, which is absurd. So long as it is reasonable to expect compliance, i.e. so long as the requirements of the theory do not violate the proviso ‘ought implies can’, the fact of actual non-compliance tells us nothing about the adequacy of the theory itself.

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 so, for such a charge presupposes a set of 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”. 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. The role of abstract principles, as O’Neill suggests, 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. Again in Kant’s words, principles “are abstracted from numerous conditions which, nonetheless, necessarily influence their practical application.” 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. 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


41 For a discussion 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 in his “Realistic and Idealistic Approaches to the Ethics of Migration”, *International Migration Review*, 30 (1) (1996), 156-165, 160ff. See also Robert Goodin’s discussion of what he calls the “accessibility critique” of ideals in “Political Ideals and Political Practice”, 39-40.

42 Immanuel Kant, “On the Common Saying: This May be True in Theory, but it does not Apply in Practice”, in Hans Reiss ed. *Kant’s Political Writings* (Cambridge: Cambridge University Press), 61-92, 61.


44 See e.g. the complex argument through which Sarah Williams Holtman shows how Kantian ideal theory might offer guidance in establishing the legitimacy of actual political decisions in her “Kant, Ideal Theory and the Justice of Exclusionary Zoning”.

45 Kant, “On the Common Saying: This May be True in Theory, but it does not Apply in Practice”, 61.
in mind. Having discussed two implausible versions of the guidance objection, I now come to consider a third, more theoretically challenging, formulation.

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”.\(^{46}\) The irrelevance charge can be succinctly articulated 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 strongly 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. Similarly, 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.\(^{48}\) Similar observations apply to cases where ideal theory designs principles of social justice. As Annette Baier protests, “descriptions of ideal societies imply *nothing* about how to act in this society”.\(^{49}\)

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*.\(^{50}\) 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. For instance, referring to the specific non-ideal case of racial discrimination, Thomas McCarthy argues: “[T]here 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.”\(^{51}\)

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 has warned about the risk of ideal theorising being *ideological*. By focusing on the design of a perfectly just social order, he argues, ideal theories obscure, rather than illuminate, existing unjust power-relations, such as those based on gender and race. By abstracting away from past and

---


\(^{47}\) This difficulty is discussed at length in Michael Phillips, “Reflections on the Transition from Ideal to Non-Ideal Theory”, *Noûs* 19 (1985), 551-570.

\(^{48}\) For similar examples see Phillips, “Reflections on the Transition from Ideal to Non-Ideal Theory”, 560-61.

\(^{49}\) Baier, “Theory and Reflective Practices”, 210, emphasis added.

\(^{50}\) Often GC2 and GC3 are not distinguished by critics. However, as I noticed, only the latter is theoretically problematic.

present injustices, so the claim goes, ideal theories surreptitiously contribute to their perpetuation, in this way defeating their own purpose: that of offering conceptual tools for criticising injustice.\textsuperscript{52}

The third interpretation of the guidance critique (GC3), 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 Rawls’s approach would have a lot to worry about. Do they? My answer is a moderate ‘yes and no’. To substantiate it, I shall take up and briefly discuss two instances of supposedly misleading idealisation, one drawn from Rawls’s domestic theory of justice (specifically, its alleged incapacity plausibly to deal with gender-based discrimination) the other from \textit{The Law of Peoples} (specifically, its alleged capacity plausibly to deal with the question of global economic justice), and consider whether GC3 successfully applies to them. I shall argue that, whilst GC3 fails in the case of gender-based discrimination, it does succeed when applied to \textit{The Law of Peoples}. The lesson Rawlsian theorists should learn from this, perhaps surprising, conclusion, will be explained in the final section of the paper.


Questions of racial and sexual discrimination are well-known both for their actual political urgency and for falling outside the scope of ‘justice as fairness’. Rawls is ready to acknowledge this omission, but he also points out (in line with my response to GC2) that

\begin{quote}

an omission is not as such a fault, either in that 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. Justice as fairness, and other liberal conceptions like it, would certainly be seriously defective should they lack the resources to articulate the political values essential to justify the legal and social institutions needed to secure equality for women and minorities.\textsuperscript{53}
\end{quote}

GC3, however, claims precisely that ideal theory \textit{lacks} those resources. Its omissions are not innocuous: 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 the claim warranted? Focusing on the specific case of gender injustice, I shall argue that it is not.\textsuperscript{54}

\textsuperscript{53} JFR, 66.
\textsuperscript{54} I think a similar argument could also be made in relation to racial discrimination and the justification of affirmative action policies, but here I shall limit myself to the case of gender injustice. Given present purposes, I do not see such a limitation as

12
The issue of the relation of Rawls’s theory to sexual injustice is widely discussed by feminist theorists and particularly by Susan Okin.\footnote{Susan Moller Okin, \textit{Justice, Gender and the Family} (New York: Basic Books, 1989), esp. 89-109.} She acknowledges that, even though Rawls does not directly discuss gender justice in his theory, the theory itself contains valuable resources to address it. Interestingly, in \textit{Justice as Fairness: A Restatement}, Rawls briefly takes up the issue of the justice of the family, where he recognises that, since it is the fundamental unit of reproduction of society, the family is to be regarded as a crucial constituent of the basic structure.\footnote{JFR, 162-168.} He also adds that his two principles apply to the family in a purely indirect fashion. Rather than positively establishing standards by which family life should be organized – this would in fact be an unacceptable violation of citizens’ private sphere – they set constraints on the forms this institution may legitimately take, ruling out those which would be incompatible with the achievement of political justice – i.e. that would violate the freedom and equality of citizens (both men and women).\footnote{JFR, 164.} That is, if a certain way of arranging the internal life of the family, of distributing benefits and burdens between husband and wife, were such as to result in an infringement of a member’s equal rights and opportunities as a citizen, political measures (in the form of laws, incentives, creation of new infrastructures etc.) to correct this form of social oppression would be justified.

For example, following Okin’s own proposal, Rawls suggests that if we were to find out that one of the major causes of women’s social disadvantage was the lack of material recognition of their contribution to society’s continued existence in the form of raising children and taking care of the family, then such a contribution should be acknowledged by promulgating laws that entitle them to half of their husbands’ earnings during the marriage.\footnote{JFR, 167.} From the viewpoint of society, by taking care of children, and ensuring that they become fully capable citizens, women’s activity is just as valuable as that of their companions, and therefore deserves acknowledgement. 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.\footnote{The later Rawls 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”, JFR, 12. See also Thomas W. Pogge, “On the Site of Distributive Justice: Reflections on Cohen and Murphy”, \textit{Philosophy and Public Affairs}, (29) (2) (2000), 137-169, esp. 165.} 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 \textit{judgements} which cannot be deduced from problematic: to show that ideal theory \textit{per se} is not doomed, I hope it is sufficient to show that it can resist GC3 at least in one respect. 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 theTopics”, in R. Goodin and P. Pettit (eds.), \textit{Contemporary Political Philosophy: An Anthology} (Oxford: Blackwell, 1997), 579-604.

\footnote{Susan Moller Okin, \textit{Justice, Gender and the Family} (New York: Basic Books, 1989), esp. 89-109.} \footnote{JFR, 162-168.} \footnote{JFR, 164.} \footnote{JFR, 167.} \footnote{The later Rawls 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”, JFR, 12. See also Thomas W. Pogge, “On the Site of Distributive Justice: Reflections on Cohen and Murphy”, \textit{Philosophy and Public Affairs}, (29) (2) (2000), 137-169, esp. 165.}
the theory itself.\textsuperscript{60} 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”\textsuperscript{61} and therefore require “an informed examination of institutional effects”\textsuperscript{62} in order to be detected and remedied. What the theory can help us do, is identify and evaluate the relative importance of the ‘things to consider’, in this way guiding us in reaching a decision. The way in which the values of liberty and equality are articulated and ‘balanced’ within justice as fairness – i.e. priority of basic liberties, fair equality of opportunity and difference principle – is meant to help us achieve exactly this. As Rawls says, “the idea of a well-ordered society should … provide some guidance in thinking about nonideal theory, and so about difficult cases of how to deal with existing injustices. It should also help to clarify the goal of reform and to identify which wrongs are more grievous and hence more urgent to correct.”\textsuperscript{63}

This being so, on the specific issue of gender, Rawls’s ideal theory may certainly be said to be incomplete (indeed, how could it be otherwise?), but without being irrelevant, or counterproductive. The foregoing, admittedly sketchy, discussion has tried to show that the theory offers a plausible conceptual apparatus from within which to capture and, possibly, remedy, the forms of gender inequality marking contemporary liberal democracies. Of course, in so saying I am not suggesting that Rawls might not be wrong on some other count, or that there might not be a theory of justice that is better than his in dealing with the question of justice and gender. Recall that what I am scrutinising here is not whether his particular ideal conception of justice is sound, but rather whether such a conception is doomed to be unsound because, like any form of ideal theorising, it must succumb to GC3. In my brief discussion so far, I have attempted to show that it is not so doomed. 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, even though GC3 has been explicitly raised against Rawls’s domestic theory of justice, it proves more successful if addressed to \textit{The Law of Peoples}.

\section*{5. Assessing the Guidance Critique (II): The Law of Peoples}

\textit{The Law of Peoples} articulates a “conception of right and justice that applies to the principles and norms of international law and practice”,\textsuperscript{64} and particularly to the “foreign policy of a reasonably just liberal

\begin{flushright}
\textsuperscript{60} Actually, 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.” JFR, 166. Here I am making the weaker claim that this is not for political philosophy alone to decide.
\textsuperscript{61} TJ, 327, emphasis added.
\textsuperscript{62} TJ, 327. In particular, Rawls thinks that violations of the difference principle are the most difficult injustices to detect, whilst violations of the first principle and “blatant violations” of the second part of the principle of fair equality of opportunity are more likely to be “obvious to all”. TJ, 326-27.
\textsuperscript{63} JFR, 13.
\textsuperscript{64} LP, 3.
\end{flushright}
people”.

Instead of being concerned with the inward dimension of liberal societies’ exercise of political power, the book focuses on its outward expression, and offers a list of eight principles for assessing its legitimate exercise:

1. Peoples are free and independent, and their freedom and independence are to be respected by other peoples.
2. Peoples are to observe treaties and undertakings.
3. Peoples are equal and are parties to the agreements that bind them.
4. Peoples are to observe a duty of non-intervention.
5. Peoples have the right of self-defense but no right to instigate war for reasons other than self-defense.
6. Peoples are to honor human rights.
7. Peoples are to observe certain specified restrictions in the conduct of war.
8. Peoples have a duty to assist other peoples living under unfavourable conditions that prevent their having a just or decent political and social regime.

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 (principles 7, 8). The latter are societies that do not meet Rawls’s international standards because they are internally oppressive or externally aggressive “outlaw states”, or otherwise disadvantaged “burdened societies”.

With the structure of Rawls’s international theory clear in mind, we can now proceed to evaluate the force of GC3 asking: 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 hardly any 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, is silent about global environmental justice and the regulation of transnational trade, and devotes no space to issues such as historical reparations for past injustices and migration policies. However, it should by now be clear that, by themselves, such omissions cannot suffice to discredit the

---

65 LP, 10, emphasis added.
66 As Rawls says, as a realistic utopia the Law of Peoples “sets limits to the reasonable exercise of power”, LP, 6 n.7.
67 LP, 37.
68 LP, 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.” LP, 106.
theoretical plausibility of Rawls’s theory – as we know, GC2 is correct but unproblematic. The point is, rather, whether in spite of such omissions The Law of Peoples can offer a plausible theoretical framework from within which to reason about what existing liberal societies ought to do in order to exercise their ‘international’ political power legitimately, and whether they are actually doing so.

Interestingly, Rawls himself takes a step in this direction, devoting part of his monograph to the discussion of principles 7 and 8, which are manifestly principles designed for non-ideal circumstances. 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”71 – liberal societies should act towards either unjust, outlaw states, or unfortunate, burdened societies. If the world were in fact 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 take on international distributive justice distorts the reality of contemporary international relations. As argued by Thomas Pogge, to see this one only has to think about past colonial domination and the perverse incentives currently generated by internationally recognised rules, such as the international borrowing and resource privileges. Privileges such as entitling whoever is in effective control of a country’s territory – i.e. independently of their legitimacy – freely to borrow and use resources in the country’s name, incentivise corrupt elites and officials to seize power in already disadvantaged societies, rendering the goal of development ever more remote. Since contemporary liberal democracies recognise, comply with, and profit from these rules, they indirectly contribute to the plight of burdened societies.72

In light of this, urging existing liberal states to ‘assist’ burdened societies whilst they have contributed and still contribute to the perpetuation of these societies’ precarious political and economic conditions, is a mischaracterisation of the moral relationship in question: 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. For even if contemporary liberal societies are far from well-ordered, The Law of Peoples generates the impression that they are, in this way obscuring, rather than illuminating, 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.73

71 LP, 89.
73 Close textual analysis reveals that Rawls is not unaware that existing liberal states are far less well-ordered than ‘his peoples’. For instance, he observes, though only in passing, that “a liberal society cannot justly require its citizens to fight in order to gain economic wealth and to acquire natural resources, much less to win power and empire. (When a society pursues these interests it no longer honors the Law of Peoples, and it becomes an outlaw state).” LP, 91. He then adds, in a footnote, that “so-called liberal societies sometimes do this, but that only shows they may act wrongly” LP, 91 n.3. True, it
Unlike the example of gender injustice offered earlier in the paper, that of global economic justice proves that there is some truth to GC3. Once we lift the full-compliance condition and consider the non-ideal circumstances marking the world today, Rawls’s principles of international justice prove unable to capture what is deeply morally problematic (in fact unjust) about the ways in which contemporary democratic societies exercise their power beyond borders.

This disanalogy between our conclusions with respect to Rawls’s ideal domestic and international theories of justice requires explanation. A helpful way to make sense of it, I suggest, is by going back to the observations advanced in section 2 regarding the unavoidability of selecting which facts to keep and which ones to bracket off when engaging in normative theorising. Given that his aim is to define the reasonable limits to liberals’ exercise of political power, we may say that, in The Law of Peoples, Rawls makes the wrong selection. Whilst, in A Theory of Justice, he seems to be aware of the channels through which power is exercised within society – i.e. through what he calls ‘the basic structure’, namely the complex set of institutions and public rules that govern the allocation of benefits and burdens among citizens – he misses it when it comes to The Law of Peoples. Rawls’s monograph asks how almost self-contained peoples ought to behave towards one another, on the assumption that, since such peoples are well-ordered, a just (fair) background is in place. However, such a background simply does not exist, and it is precisely its absence that gives rise to the question of international justice in the first place.

In ideal theory, where liberal societies are all 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 that are unfavourable to them. This is certainly a highly desirable social ideal, yet one that is very far, too far, from the status quo, which is why, in the case of The Law of Peoples, the transition from ideal theory to non-ideal circumstances cannot be successfully carried through. A successful transition would in fact require abandoning the assumption (idealisation) that liberal societies are well-ordered and almost self-contained. Yet such assumptions constitute what we might think of as the ‘basic structure’ of Rawls’s international theory, the framework from which the theory invites us to look at the world. To alter such a framework, is to dispense with the theory altogether. The only alternative to such disruptive alterations, is to ‘apply’ The Law of Peoples to non-ideal circumstances in the way Rawls himself does, and continue to depict existing liberal societies as almost self-sufficient, well-ordered peoples, when in real international politics they take
advantage of their superior bargaining position and negotiate ‘terms of cooperation’ that are particularly ‘burdensome’ to less powerful nations – terms from which such nations cannot realistically subtract themselves.\(^{76}\)

In short, being designed to impose limits on the exercise of political power, and virtually ignoring the ways in which it is actually exercised, *The Law of Peoples* represents a bad example of ideal theory, one that is doomed to be irrelevant to the very problem it sets out to solve. The idealisations Rawls builds in the theory defeat the purpose of the theory itself, for once we realise that the world does not work as the theory implicitly assumes – i.e. once we realise that societies are not self-contained, and well-ordered in their international behaviour, in fact they are not peoples in Rawls’s sense – in order to criticise and evaluate the current international scenario, we seem to have to resort to another theory altogether. Instead of helping us make sense of what international justice requires today, *The Law of Peoples* is more apt as a theory of justice for what Allen Buchanan calls a “Vanished Westphalian World”.\(^{77}\)

### Conclusion

We started our 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 two examples drawn from ideal theories which both aspire to be action-guiding: Rawls’s ‘justice as fairness’ and the Law of Peoples. Interestingly, the two cases have yielded different conclusions. With ‘justice as fairness’, the gap between ideal and non-ideal theory has proven bridgeable (at least in relation to the case of gender injustice), whilst this has not been the case for the Law of Peoples. This confirms the suggestion, advanced in section 2, that there is nothing wrong with ideal theory *per se*, and that some form of ideal theorising is virtually unavoidable in normative thinking.\(^{78}\)

Yet, the disanalogy between our two examples also shows that there is some merit to the claim that “reliance on idealization to the exclusion, or at least marginalization, of the actual”,\(^{79}\) may turn out to be morally problematic. The guidance critique is, therefore, not altogether misguided. Whether a certain ideal theory is objectionable depends on the appropriateness of its idealisations, given the question the theory addresses. Those who criticise ideal theory, are in fact pointing to what we might call *unwarranted* idealisations, such as those found in Rawls’s Law of Peoples. Such idealisation build into a normative

\(^{76}\) On this see, e.g., David Singh Grewal, “Network Power and Globalization”, *Ethics and International Affairs*, 17 (2) (2003), 89-98.  
\(^{77}\) Buchanan, “Rawls’s Law of Peoples: Rules for a Vanished Westphalian World”.  
\(^{78}\) Interestingly, this is implicitly acknowledged even by advocates of *non*-ideal theory such as Charles Mills, for instance when he says that “nonideal theory can and does appeal to an ideal”, in Mills, “Ideal Theory as Ideology”, 171.  
\(^{79}\) Mills, “Ideal Theory as Ideology”, 168.
theory a factually inaccurate 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.

We can therefore conclude that the paradox can be overcome. Ideal theorising is unavoidable, but this does not show that the enterprise of normative political theory is doomed. Whether an ideal theory is good or bad depends on whether its idealisations are warranted or not, and this in turn can only be evaluated through substantive argument, in light of the specific aim of the theory. The lesson to be learnt from the guidance critique is of a methodological nature, and urges theorists of justice not to lose sight on the facts, or “conditions of the world”, that are relevant to the question they are attempting to answer. In the case of Rawlsian justice, this is the way in which liberals actually exercise political power, within existing practices. Keeping the facts in sight in this way, does not entail subservience to the status quo; rather, it is crucial in order to get ‘the distance between the ideal and the real’ right, so as to produce a theory that is both critical and action-guiding.

---

80 Cf. LP, 90.