Ideal Theory and Utopia

Volker Ruitinga

Introduction

Social and political philosophy has inherited a great deal from the work of John Rawls. He has been widely credited with reviving social theory in the 20th century. Amongst his many contributions is an intuitively simple concept which has sparked controversy and debate in the post-Rawlsian literature on theories of justice: that of Ideal Theory, as opposed to non-Ideal Theory. What seems like a simple distinction has become the subject of debate for two important reasons: firstly, Rawls is not particularly concerned with the distinction and leaves it ill-defined. To illustrate: in his work *A Theory of Justice* the words 'ideal' and 'theory' appear in conjunction a mere five times. Yet, the Ideal / non-Ideal distinction strikes at the heart of social theory. If we take the basic goal of social theory to be advancing the betterment of society, many scholars agree that some ideal conception of society is of value to this end (see Bloch, 1986; Davis, 1981; Levitas, 2011). Secondly, even if Rawls had provided a clear definition of the distinction between Ideal and non-Ideal Theory, controversy surrounding the object of the distinction would remain.

Rawls defined his own work as Ideal Theory. Consequently, the literature on social and political philosophy has seen several attempts to categorize theories of justice based on the Ideal/non-Ideal distinction (e.g. Stemplowska, 2008); an obvious problem with this endeavour being that there is no consensus on what the definition of this distinction may be. Therefore, many of these efforts have only succeeded in highlighting certain difficulties surrounding the debate. This goes to show that both the concept of Ideal Theory and therewith the supposedly 'ideal theories' themselves are the source of much disagreement. Thus, we must examine the nature of Ideal Theory, its contours and defining characteristics, before

we can constructively discuss the consequences of this distinction on the theories in question.

I wish to start by suggesting that all engagement in ideal theorizing has at its source some conception of an ideal society. Outside the post-Rawlsian literature on theories of justice this concept has often been referred to as *Utopia*. However, in the literature on Rawlsian Ideal Theory there is a marked absence of utopian terminology. One of the reasons for this absence may be that scholars are wary of using the term because it is itself the subject of much controversy. Another reason may be that the term is traditionally used to refer to more than issues of justice alone. However, although these are defensible grounds for introducing new terminology, unique to theories of justice, ignoring utopian literature results in a number of missed opportunities. Firstly, there is undoubtedly significant overlap between Utopian Theory and Ideal Theory. Ideal theories of social justice can potentially be categorized as a subset of Utopian Theory, alongside for example Utopian- Moral, Political and Legal Theory. Secondly, owing to this overlap, examining the literature on Utopian Theory may reveal significant parallels between the two debates. Thirdly, and most importantly, integrating the concept of utopia into the definition of Ideal Theory could serve to clarify the concept. This will, in turn, lead to a more fruitful use of the distinction.

In light of this observation the aim of this paper is twofold. Firstly, to suggest a new way of understanding Ideal Theory, specifically with reference to utopian literature, in order to illustrate the benefit of introducing the concept Utopia to the existing literature on Ideal Theory. To this end I will re-examine the Rawlsian definition of Ideal Theory and briefly sketch the similarities between Ideal Theory and Utopia. I aim to show that Ideal

Theories of Justice should be regarded as Utopian. Secondly, I will look at two articles that discuss the relevance of Ideal Theory to see if my proposed amendments aid the defence of Ideal Theory against criticism, as I consider Ideal Theory and Utopia to be essential to social theory.[1]

Part I

In order to compare utopianism with Rawlsian Ideal Theory, a brief introduction to Utopia is required. In 1516 Sir Thomas More gave the name Utopia to his proposed ideal society. At that time the word was a neologism, or more accurately a lexical neologism, meaning that it named a new concept or synthesized pre-existing ones (Vieira, 2010: 3). It combined the Greek words ou (οὐ), meaning *no*, and topos (τόπος), meaning *place*, adding the suffice '*ia*' indicating *a* place. More originally intended to use the word Nusquama, nusquam being Latin for 'nowhere' or 'never'. However, he chose Utopia, for when spoken in English both Outopia and Eutopia are phonetically alike, the latter meaning *good* place. This eloquently captured the ambiguity of any such imagined society, the unceasing tension between the ideal and the unreachable.

Utopian *Theory* is essentially concerned with conceptualizing the 'ideal commonwealth', which inherently expresses a 'psychological aspiration of hope for a better state of existence in this life or elsewhere, notably in the form of the quest for "community"' (Claeys, 2010: xi). Often this 'theorising' has taken the form of literature, as in Thomas More's *Utopia*. But Utopian Theory encompasses much more than enticing storytelling alone. Karl Mannheim famously wrote on the relationship between Utopia and ideology (Mannheim, 1954), Bloch on Utopia and hope (Bloch, 1985), Goodwin on Utopia and politics (Goodwin, 2009), and so on. Utopian Theory encompasses many aspects of social theory, because Utopia is so fundamental to our thinking about society. With this in mind, we turn to the examination of Rawlsian Ideal Theory.

Rawlsian Ideal Theory and Utopia

There are many interpretations of Ideal Theory. For example, some scholars (erroneously) equate Ideal Theory solely with the condition of full-compliance (to the conditions prescribed for society), one problem with this equation being that full compliance 'may also hold for principles of justice which do not lead to a just society' (Robeyns, 2008: 3). Moreover, nowhere does Rawls say that it is a sufficient condition for Ideal Theory. However, this does beg the question: what conditions are necessary for an Ideal Theory according to Rawls? In order to answer this we must highlight a few key passages regarding Ideal Theory from *A Theory of Justice*.

To start, Rawls limits the scope of his inquiries in several respects. For instance, he is concerned only with instances of justice, for 'justice is the first virtue of social institutions' (Rawls, 1999: 3). A further limitation is best summarized by the following: 'I shall be satisfied if it is possible to formulate a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies' (Rawls, 1999: 7). Here Rawls emphasises the need for theoretical simplification by stating that his ideal society is both closed and isolated. Rawls also writes: '[...] I consider primarily what I call strict compliance as opposed to partial compliance theory' (Rawls, 1999: 7). Crucially, he goes on to say that *partial* compliance theory 'studies the principles that govern how we deal with injustice' (Rawls, 1999: 7), meaning that when there is full-compliance to the hypothesized principles of the theory of justice there can be no instances of injustice.

So far we know that Ideal Theory is concerned with providing a theory of justice for an isolated society, whose members act in full compliance to the proposed principles of justice, resulting in a situation where there are no instances of injustice. This leaves a very narrow definition of Ideal Theory, as any theory concerning partial compliance or indeed one that results in any instances of injustice (resulting from proposed institutional policy) would not be considered an Ideal Theory. Rawls goes on to defend this view by stating: 'The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. [...] At least, I shall assume that a deeper understanding can be gained in no other way, and that the nature and aims of a perfectly

just society is the fundamental part of the theory of justice' (Rawls, 1999: 8). Hence, Rawls claims that there can be no complete non-Ideal Theory without Ideal Theory.

Rawls comes closest to a definition of Ideal Theory in the following passage:

> 'The intuitive idea is to split the theory of justice into two parts. The first or ideal part assumes strict compliance and works out the principles that characterize a well-ordered society under favorable circumstances. It develops the conception of a perfectly just basic structure and the corresponding duties and obligations of persons under the fixed constraints of human life' (Rawls, 1999: 216).

Here, Rawls points out that the endeavour of designing an Ideal Theory requires certain theoretical limitations. One of which is that it requires the paradoxical assumption of favourable circumstances whilst accepting the fixed constraints of human life; (paradoxical because the constraints of human life are often unfavourable to creating an ideal society).

At this point it must be noted that many of these same conditions and limitations also hold for conceptions of Utopia. Utopian Theory often envisions an isolated society, under favourable conditions, where there can be no injustice if its members comply to the societal ideals. This goes to show that the definition of Rawlsian Ideal Theory arguably holds as a viable, although simplified, definition of Utopia. We can then define Ideal Theory to be: *a system of principles that, when fully-complied to by all members of society, results in a Utopia of social justice.*

Ideal Theory and Utopianism

Against this backdrop, a number of significant similarities between Utopian Theory and Ideal Theory become apparent. For instance, both Utopian Theory and Ideal Theory envision some significantly improved version of society. Furthermore, neither theory is primarily concerned with explicating the transition to this ideal from our world.

But there are other, more subtle parallels, for example: both theories

attach some specific value to their imagined place. In Eutopia this is the *good*, in Ecotopia, Vegatopia and Technotopia their overriding values are clear. In this vein, Rawslian Ideal Theory is primarily concerned with *justice*. Additionally, they share the same potential to guide human progress by presenting a well-argued example. Conversely, both Utopian Theory and Ideal Theory are subject to many of the same criticisms: their ideas are said to be unreachable fantasies and to pursue them is a waste of time. Furthermore, it is no coincidence that Sir Thomas More chose to situate Utopia on an island, extremely well-guarded from the outside world (More, 2007: 33), whilst John Rawls strives to formulate 'a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies' (Rawls, 1999: 7).[2]

Both the work of John Rawls and many Utopian scholars share the belief that the social transition toward some ideal requires a conception of that ideal. But, the work of John Rawls differs from the Utopian theory of Mannheim, Bloch and Goodwin in that it does not explicitly discuss the relationship of its *object* (in Rawls: social justice) to Utopia. However, if the concept of Utopia can be integrated into Ideal Theory it becomes much more comparable to the works mentioned above as utopian. With this, the compatibility of the different works on ideal societies can be examined, opening up the possibility of a wider, cross-discipline (or cross-object) account of the theory of ideal societies. Thus, as we have seen these scholars profess the importance of the concept Utopia to their respective theories, we can suggest that Utopia could be of similar importance for (ideal) theories of justice. At the very least it may serve to highlight the parallels that are currently overlooked by ignoring Utopia.

Now we can summarize the initial benefits of viewing Ideal Theory as a Utopian Theory. Firstly, the inclusion of the concept of Utopia serves to remind us that Ideal Theory should not be limited to theories of justice alone. Utopia's most often advocate some specific virtue(s), hence; there can also be Ideal Theories of Morality, Happiness, Freedom, etc.[3] Consequently, we should refer to Rawlsian Ideal Theory as an Ideal Theory of Justice. This frees the term Ideal Theory up to be used with reference to other disciplines. Secondly, the incorporation of Utopia emphasises

that the Rawlsian definition of Ideal Theory is a narrow one; Ideal Theory should do more than achieve a reasonably good society, relatively free from injustice. The preceding two reasons demonstrate the most important benefit of introducing Utopia to Ideal Theory: it serves as conceptual clarification. Furthermore, referring to a Utopia as the product of Ideal Theory, and thus of Ideal Theories of Justice, underlines the fact that in Rawlsian Ideal Theory: (1) some favourable conditions are assumed (this is arguably true of all Utopia's), (2) full-compliance is a condition and (3) that there can be no instances of societal injustice. This then makes Ideal Theories of Justice utopian. Additionally, Rawlsian Ideal Theory can benefit from the warning of history often associated with Utopia in our post-communist world.

Thus far I have re-examined Rawlsian Ideal Theory, sketched its parallels with Utopian Theory and concluded that the use of Utopia benefits its conceptual clarity. Now we proceed to the second aim of this paper, which is to examine the consequences of this new point of view on two articles from contemporary literature regarding Rawlsian Ideal Theory. The first article defends Ideal Theory, whilst the second stands in opposition.

Part II

Zofia Stemplowska: *What's Ideal About Ideal Theory?*

Zofia Stemplowska opens her paper by remarking that Ideal Theories 'share a common characteristic: much of what they say offers no immediate or workable solution to any of the problems our societies face' (Stemplowska, 2008: 319). She does not, however, consider this a fatal flaw and sets out to defend these theories by contending that the debate regarding Ideal and non-Ideal theories can be productive, if they are not treated as rival approaches to political theory. To this end she offers her own definition of Ideal Theory by examining the structure of normative theory. Initially she writes:

'One crucial difference between various normative theories concerns whether they offer viable recommendations, where by viable recommendations I mean recommendations that are both achievable and desirable' (Stemplowska, 2008: 324).

Stemplowska calls these 'AD-recommendations' and believes that it is the absence of these recommendations that is crucial in separating Ideal from non-Ideal Theories. Consequently, Stemplowska defines non-ideal theory as 'theory that issues AD-recommendations, and ideal theory as theory that does not' (Stemplowska, 2008: 324). With this definition she goes on to say that normative theories may lack these AD-recommendations for different reasons. Firstly, they may offer recommendations that cannot be considered AD-recommendations, and secondly they may not aim at offering any such recommendations at all. She claims that the latter serve only to clarify our understanding of certain values and principles, and can therefore not be objectionable. Of the former, she proceeds to identify several further sub-categories. Ignoring what she calls 'bad theories', Stemplowska identifies '(a) theories that fail to issue AD-recommendations because they ignore the fact of non-marginal noncompliance, and (b) theories that fail to issue AD-recommendations because, even with full compliance, there is no solution to the problem for which recommendations are sought' (Stemplowska, 2008: 331).

In what follows, Stemplowska defends theories that do not offer AD-recommendations, concluding that they are nevertheless indispensable to normative theory. Furthermore, she claims that accepting her definition of the distinction between Ideal and non-Ideal Theory allows us to see that complex normative theories are likely to contain both Ideal and non-Ideal aspects.

Although Stemplowska defends Ideal Theory, her contention that the identification of AD-recommendations best resolves what is at stake in the debate between Ideal and non-Ideal Theories fails to be convincing. The main problem with her approach is that she offers a negative definition of Ideal Theory, meaning she defines Ideal Theory by what it is not. She identifies a characteristic that is most often associated with non-Ideal Theory and then attempts to define and categorize Ideal Theories by virtue

of the absence of this characteristic. This results in the need to categorise different Ideal Theories according to why they do not meet the criterion of supplying AD-recommendations. With each of these additional categories come further issues of definition. The obvious problem with this approach is that it is unclear how these divisions are to be made.

Furthermore, Stemplowska approaches the definition of the distinction between Ideal and non-Ideal Theory from the bottom-up. Meaning, she takes a characteristic of the majority and defines the minority by its absence. Consequently, her definition of non-Ideal Theory is too broad. Undoubtedly there is a much greater body of non-Ideal Theory, but using a common feature of the majority as the basis for a negative definition of the minority may lead to several undesirable consequences: (1) Ideal Theories may falsely be labelled non-Ideal; some of Rawls' recommendations are arguably both achievable and desirable, for example, and (2) non-Ideal theories may be labelled ideal, as their output could be argued to be both unachievable and/or undesirable. It would be more fruitful to characterise Ideal Theories by virtue of some unique feature, as opposed to the absence of a common feature. In other words, what is required is a top-down approach. The amended definition of Ideal Theory with reference to Utopia is an example of this top-down method for identifying Ideal Theories. Defining Ideal Theory as a system of principles that, when fully-complied to by all members of society, results in a Utopia of social justice, would preclude the need for the problematic sub-categories of Ideal Theory described by Stemplowska. The image of the top-down approach thus captures the Rawlsian idea that some Ideal Theory is required primary to non-Ideal theorising. Moreover, the addition of Utopia to the definition of Ideal Theory also serves as a *necessary* condition of Ideal Theory, as any non-Ideal Theory cannot result in Utopia.

Although Stemplowska's proposal faces problems that are likely insurmountable, she succeeds in highlighting a poignant difference between Ideal and non-Ideal theory in general. Unfortunately, the resulting attempt to distinguish between Ideal and non-Ideal Theory based on this difference is highly problematic and therefore not useful as a tool for differentiation. Besides lacking the theoretical virtue of being narrow, the problems affecting Stemplowska's proposal suggest that an approach singling out a common positive feature of Ideal Theory is preferable.

Charles W. Mills: *"Ideal Theory" as Ideology*

Having examined an article that accepts the need for Ideal Theory, it is important to discuss another that rejects it. This will show us if the new understanding of Ideal Theory can survive established criticisms. Charles W. Mills proves himself to be a vocal opponent of Ideal Theory in his *"Ideal Theory" as Ideology*. His article is nothing short of an all-out attack on Ideal Theorizing. He sets out not only to discredit Ideal Theory, but to prove that non-Ideal Theory is superior in every way; going so far as to say that even the act of engaging in Ideal theorizing perpetuates the non-ideal (Mills, 2005: 182). Mills proposes that only Non-Ideal theorizing can offer solutions to the non-ideal. To serve his ends, Mills employs issues such as gender and race inequalities to demonstrate the need for non-Ideal Theory. Throughout his article, Mills offers possible definitions of Ideal Theory and argues why these do not hold. I wish to show that on two occasions Mills mistakenly dismisses Ideal Theories.

Mills begins by distinguishing different types of theorizing of which the most important, in this context, are ideal-as-idealized and ideal-as-descriptive. The former referring to an idealized model of what some ideal P should be like, the latter being a somewhat idealized or abstracted model of how P actually works. Mills then builds on these to define Ideal Theory, he writes: 'What distinguishes ideal theory is the reliance on idealization to the exclusion, or at least the marginalization, of the actual' (Mills, 2005: 168). This is of course not a strict definition of Ideal Theory for it is difficult to determine the extent of reliance on idealization, let alone the marginalization of the actual or even what the actual may be. However, Mills does go on to specify Ideal Theory further: 'ideal theory either tacitly represents the actual as a simple deviation from the ideal, not worth theorizing in its own right, or claims that starting from the ideal is at least the best way of realising it' (Mills, 2005: 168). Firstly, I do not believe that any substantive Ideal Theory tacitly represents the actual as a *simple* deviation from the ideal. Secondly, none hold that this is not worth theorising in its own right. According to Mills then, Ideal Theories must then claim that

starting from the ideal is the best way of realizing the ideal, a claim which he believes to be false.

Mills defends this claim by quoting John Rawls: 'The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems' (Rawls, 1999: 8). Mills mistakenly equates his own claim, that in the view of Ideal theorists 'starting from the ideal is the best way of realizing it', with Rawls' statement. He compounds this mistake by later adding: 'the argument has to be, as in the quote from Rawls above, that this is the *best* way of doing normative theory, better than all the other contenders' (Mills, 2005: 171 – italics in original). To start, Rawls certainly does not claim that Ideal Theorizing is the best way to realize the ideal, only that it is a necessary component of this process. Additionally, he makes no claim to have discovered 'the *best* way of doing normative theory'. What Mills points out at best is that most, if not all, Ideal theorists claim that Ideal Theory is a necessary step toward realizing the ideal; not the best nor the only.

Amongst Mills' many criticisms there are two in particular that illuminate the error in his dismissal of Ideal Theory. In a lengthy section on *The Vices of Ideal Theory*, Mills wishes to 'quickly clear away some of the ambiguities and verbal confusions that might mistakenly lead one to support ideal theory' (Mills, 2005: 170). Of these 'verbal confusions' the first conception of Ideal Theory that is mistakenly dismissed by Mills is that of Ideal Theory being 'just a model' (Mills, 2005: 171). It can be defended that Ideal Theories are just that: theories. As has been said, Rawls was satisfied with the possibility of formulating a reasonable conception of justice for the basic structure of society (Rawls, 1999: 7). On the other hand, far from this being *just* a theory, any reasonably successful attempt to model an ideal society based on a set of principles would be no small feat, and would be invaluable to the study of political and social theory in philosophy. Moreover, an Ideal Theory need not say anything about the non-Ideal, nor offer any value judgements or achievable and desirable recommendations for that matter. It could function as 'just a model', placing the burden of implementation on non-Ideal Theories. Ideal Theory may serve only to demonstrate or test the compatibility of certain ideals proposed in a theory. If these professed ideals prove to be compatible, the

Ideal Theory would be the blueprint for Utopia.

The second misconception reads 'Nor does the simple appeal *to* an ideal (say, the picture of an ideally just society) necessarily make the theory ideal theory, since nonideal theory can and does appeal to an ideal also' (Mills, 2005: 171 – italics in original). However, I would argue the exact opposite. Appealing to a 'picture of an ideal society', or Utopia, *does* necessarily make the theory Ideal Theory. The fact that a paradoxical assumption of favourable conditions is made, the fact that there must be full-compliance to the prescribed ideals, the fact that there can be no instances of injustice and the fact that these conditions constitute the ideal state, or Utopia, make such theories Ideal Theories. The inclusion of Utopia only serves to *underline* that these conditions are part of Ideal Theory. Moreover, the appeal of any non-Ideal Theory to some ideal would require some conception of that ideal, which in turn requires some theory of said ideal. It is at this point that non-Ideal Theory necessarily appeals to Ideal Theory. In other words, Mills was wrong to reject Ideal Theory on the basis of these arguments. Moreover, we can better understand why Mills fails to reject Ideal Theory if we refer to Ideal Theory with reference to Utopia.

In conclusion, the addition of Utopia to the conception of Ideal Theory helps save it from the Mills' criticism by highlighting where he falsely dismisses Ideal Theory. Mills does not succeed in relegating Ideal Theory to a sub-par status in social and political philosophy. In fact, if Mills were to advocate some version of Ideal Theory it would be one incorporating Utopia, for with Utopia come the many warnings of history against blindly implementing ideology. He would fervently endorse the historic dimension that Utopia brings to Ideal Theory as he himself so often appeals to it.

Conclusion

In this paper I initially re-examined the Rawlsian definition of Ideal Theory and explored the relationship between Ideal Theory and Utopian Theory. After finding structural similarities, I introduced the concept of Utopia to Ideal Theory of Justice, resulting in a new definition of Ideal Theory. A complete Ideal Theory of Justice would be: a system of principles that, when fully-complied to by all members of society, results in a Utopia of

social justice. I found support for this new definition by examining Zofia Stemplowska's article on Ideal Theory, which demonstrated the need for a positive definition of Ideal Theory, meaning that it must identify some characteristic of Ideal Theory that non-Ideal theories lack, such as a conception of Utopia, as opposed to being labelled Ideal by virtue of lacking some feature that non-Ideal theories share. Additionally, the new conception of Ideal Theory withstands attempts by Charles W. Mills to render all Ideal Theory irrelevant.

To conclude, I hope to have shown that a definition of Ideal Theory with reference to the ideal society Utopia, is not only possible, but desirable; for it can both clarify the debate and withstand significant criticism. With this, we may be encouraged to look more closely at the similarities between Utopianism and Ideal Theory.

Volker Ruitinga (1987) completed his Bachelor's degree in Philosophy at the Erasmus University, and is currently enrolled in the Master of Philosophy. Over the course of his studies he developed an interest in Utopia, and is currently writing his master thesis on the concept of Utopia and Ernst Bloch's 'Das Prinzip Hoffnung'.

'Ideal Theories and Utopia' was written for the completion of the mastercourse 'Contemporary Theories of Justice' by prof. dr. Ingrid Robeyns.

Literature

Bloch, E. (1985) *Das Prinzip Hoffnung*. Frankfurt am Main: Suhrkamp.

Claeys, G. (2010) 'Preface'. In G. Claeys (ed.) *The Cambridge Companion to Utopian Literature*. Cambridge: Cambridge University Press.

Davis, J (1981) *Utopia and the Ideal Society*. Cambridge: Cambridge University Press.

Goodwin, B. & Taylor, K. (2009) *The Politics of Utopia*. Bern: Peter Lang.

Levitas, R. (2011) *The Concept of Utopia*. Oxford: Peter Lang.

Manheim, K. (1954) *Ideology and Utopia*. New York: Harcourt, Brace.

Mills, C.W. (2005) '"Ideal Theory" as Ideology'. In: *Hypatia* 20, No.3.

More, T. (2007) *Utopia*. Sioux Falls: NuVision Publications.

Rawls, J. (1999) *A Theory of Justice: Revised Edition*. Harvard: Harvard University Press.

Robeyns, I. (2008) 'Ideal Theory in Theory and Practice'. In: *Social Theory and Practice*, Vol.34, No.3.

Stemplowska, Z. (2008) 'What is Ideal About Ideal Theory?'. In: *Social Theory and Practice*, Vol.34, No. 3.

Vieira, F. (2010) 'The concept of utopia'. In: G. Clayeys (ed.) *The Cambridge Companion to Utopian Literature*. Cambridge University Press.

54

Notes

1 Elaboration on the reasons for this conviction are beyond the scope of this essay but rely generally on the idea that to change the world for the better requires some conception of what this better world would be.

2 Moreover, it is no coincidence that both More and Rawls introduce their works by citing these conditions.

3 An Ideal Theory of Justice envisions a Utopia free from injustice, whereas an Ideal Theory of Freedom, for example, would conceptualize a society whose members suffer the bare minimum of constraints on their actions.