

Universal Principle of Right: Metaphysics, Politics, and Conflict Resolutions¹

SORIN BAIASU

Keele University

Email: s.baiasu@keele.ac.uk

Abstract

In spite of its dominance, there are well-known problems with Rawls's method of reflective equilibrium (MRE), as a method of justification in meta-ethics. One issue in particular has preoccupied commentators, namely, the capacity of this method to provide a convincing account of the objectivity of our moral beliefs. Call this the Lack-of-Objectivity Charge. One aim of this article is to examine the charge within the context of Rawls's later philosophy, and I claim that the lack-of-objectivity charge remains unanswered. A second aim of this article is to examine the extent to which, despite Rawls's express intention to avoid reliance on Kant's moral philosophy, supplementing Rawls's political constructivism with some Kantian elements, in particular Kant's idea of a universal principle of right, not only addresses some of the issues raised by the lack-of-objectivity charge, but also does so without compromising the ability of the Rawlsian account to accommodate the pluralism of conceptions of the good, which he takes to be a fact of modern democracies. I argue for a revised justificatory methodology, which combines Rawls's MRE and Kant's Critical Method.

Keywords: method of reflective equilibrium, transcendental arguments, Critical method, justification, objectivity, Kant, Rawls

1. Introduction

In spite of its dominance, there are well-known problems with Rawls's method of reflective equilibrium (henceforth, MRE or RE), as a method of justification in meta-ethics. One issue in particular has preoccupied commentators, namely, the capacity of this method to provide a convincing account of the objectivity of our moral beliefs.² Call this the Lack-of-

Objectivity Charge. Although defenders have not hesitated to respond to this challenge,³ the view that the MRE cannot provide an account of moral objectivity, despite Rawls's explicit intention to offer it, continues to be advanced in the literature⁴ and alternatives are being actually sought for.⁵

Rawls uses the MRE in both his early and in his later philosophy, but the relevance of the distinction between the early and the later Rawls for his view of the MRE is usually not considered in the literature.⁶ In particular, the fact that the later Rawls makes claims, which are much less ambitious philosophically, but much more engaged politically, should raise the question whether, despite what seem to be successful formulations of versions of the lack-of-objectivity charge, the MRE is not in fact sufficient for the purposes of the later Rawls. For instance, as we will see, the later Rawls distinguishes between various senses of objectivity as variously appropriate depending on the moral theory considered, and he specifically adopts a strategy of avoidance of controversial philosophical topics as part of the theory he defends, namely, political constructivism.

One aim of this article is to examine Rawls's MRE within the context of his later philosophy, and to determine the extent to which the lack-of-objectivity charge still applies. I start from what seems to be one of the strongest defences of Rawls's method, I examine Rawls's account of the objectivity of political constructivism, but I eventually conclude that the lack-of-objectivity charge remains unanswered. The shift in Rawls's later political theory towards a philosophically less ambitious account of justice was prompted in part by the realization that his early philosophy relied on a Kantian version of constructivism, which it was not plausible to expect as a ground within the context of the modern democracies' pluralism of conceptions of the good. For the purpose of a theory of justice, which is able to adjudicate upon, and reconcile conflicts between, such conceptions of the good, Rawls adopts political constructivism, a view which is non-committal to, among others, Kantian constructivism.

A second aim of this article is to examine the extent to which, despite Rawls's express intention to avoid reliance on Kant's moral philosophy, supplementing political constructivism with some Kantian elements, in particular Kant's idea of a universal principle of right, not only addresses some of the issues raised by the lack-of-objectivity charge, but also does so without compromising the ability of the Rawlsian account to accommodate the pluralism of conceptions of the good, which he takes to be a fact of modern democracies. My conclusion will be that some of the important reservations which can be expressed in relation to such a

project are less serious than they may seem, and the Rawlsian idea of a political conception of justice which can fairly reconcile conflicts between conceptions of the good can be more easily pursued when some Kantian elements are allowed back within the later Rawls's view of justice.

In section 2, I present a reading and defence of Rawls's MRE, which claims to show that the Rawlsian method is the best and the only method of moral justification. Section 3 evaluates this claim and concludes that the lack-of-objectivity charge continues to stand; moreover, in that section, I identify the element which the Rawlsian account misses, in order to offer a better account of the objectivity of our moral beliefs. Section 4 considers whether this missing element, namely, the Kantian Universal Principle of Right, could complement Rawls's political constructivism, while at the same time being as accommodating of a pluralism of conceptions of the good as Rawls's constructivism. I defend the view that a Rawlsian political constructivism enriched with Kantian elements can successfully function as a method of reconciling conflicts between conceptions of the good, even if this will require the incorporation of more metaphysical elements than Rawls would have initially bargained for. Further clarification of the revised MRE through consideration of three objections is offered in the concluding section.

2. Rawls's MRE as the Best and Only Method of Justification

According to T. M. Scanlon, Rawls's MRE,

properly understood, is in fact the best way of making up one's mind about moral matters and about many other subjects. Indeed, it is the only defensible method: apparent alternatives to it are illusory. ([2003] 2006: 149)

As my aim in the next section is critical of the MRE, it seems apposite to start from as broad and strong account of Rawls's method of justification as possible. Scanlon's reading of the MRE is in fact part of an interpretation of Rawls's account of justification in general, an account which includes two other methods of justification. Moreover, as we have just seen, he thinks that he can defend Rawls's account (for which the MRE is central) as the best and only method of moral justification.

Apart from the method of reflective equilibrium, Scanlon also examines the derivation of the principles of justice in the original position and the

idea of public reason. However, it gradually becomes clear that the method of reflective equilibrium is the fundamental method of justification, since, on Scanlon's reading, the structure of the original position is itself justified by employing the method of reflective equilibrium (Scanlon [2003] 2006: 153) and the idea of public reason, with its two significant components – the overlapping consensus and the criterion of reciprocity – can also be represented by means of the method of reflective equilibrium (Scanlon [2003] 2006: 160, 161).

In the presentation of Rawls's MRE, Scanlon begins with a distinction between two ways in which 'justification' can be understood. On the one hand, we can say that a claim is justified, when it is supported by good and sufficient reasons. On the other hand, we can say that a person is justified in making a certain claim when she makes that claim in virtue of reasons she reasonably takes to be good and sufficient. It follows that a person may be justified in holding a certain claim, although the claim itself is not justified, due to other factors, which, unbeknown to her, undermine the justificatory force of the reasons (Scanlon [2003] 2006: 140).

Now, the MRE can be presented as a three-stage process. There is first the stage at which a set of considered judgements about justice is identified. There is, then, the stage where principles are formulated which account for these considered judgements. Finally, there is the stage which includes a process of reconciling divergences between principles and considered judgements: one can abandon the considered judgements, which are in conflict with the principles, or reformulate the principles, in order to achieve a better fit with the considered judgements, or combine these two responses. The state where there is no conflict between the remaining considered judgements and the reformulated principles is called Reflective Equilibrium.

Various aspects of the method need to be clarified further. First, considered judgements are those claims which seem to be correct under conditions which are conducive to making good claims, such as being fully informed about the matter in question, thinking carefully and clearly about the relevant issues and not being subject to factors that are likely to distort judgement. Moreover, considered judgements need not be about particular cases, but can include claims at all levels of generality.

Secondly, it is not immediately clear whether the method has a specifically descriptive or normative aim. Scanlon notes that there are several

instances in Rawls's writings which suggest a descriptive aim, but a particularly interesting one is from Rawls's 'The Independence of Moral Theory' (Rawls 1974). Rawls notes there that the history of moral philosophy shows that the notion of moral truth is problematic, so he suggests suspending its consideration until a greater understanding of the various moral conceptions is achieved. (Rawls 1974: 7; Scanlon [2003] 2006: 142). As people are influenced by moral conceptions, these conceptions can be made a focus of study. The moral theorist, Rawls notes, can be seen as an observer who attempts to make explicit the structure of other people's moral conceptions and attitudes.

There is, however, a distinction in this article by Rawls, which suggests that the MRE can be read normatively. Thus, Rawls adds that, in studying one's own moral conception and attitude, one needs to distinguish between one's role as moral theorist and one's role as a person who applies a moral conception in her life. The moral theorist examines 'an aspect of human psychology, the structure of our moral sensibility'; the moral person applies a moral conception, as a correct theory about what is objectively right or wrong (Rawls 1974: 288; Scanlon [2003] 2006: 142). This second role suggests the MRE might have the role of helping us to determine what we believe about justice, that is, might help us to deliberate about normativity.

According to the descriptive reading, the MRE can be a method of justification in the second sense of 'justification' presented above: it helps us to present accurately the conception a person happens to have and the reasons for this conception she can reasonably take to be good and sufficient. Hence, the MRE can be seen as a method of presenting accurately various conceptions (say, moral conceptions or conceptions of justice), between which we can choose on the basis of some other method. This other method can also be the MRE, but according to the deliberative reading. On the deliberative reading, the MRE can help us to choose between various available conceptions and become committed to the best conception; hence on this reading, the MRE would be a method of justification in the first sense of 'justification'.

Although we may seem to deal here with two conflicting interpretations of the MRE, Scanlon notes that the two interpretations are in fact complementary and the deliberative one has priority over the descriptive reading. Thus, even if the aim of using the MRE is descriptive, the person whose considered judgements are in question must have been trying to decide what to believe and the MRE seems to be the appropriate method

for this, especially if we look at its third stage, where divergences between the principles formulated and the considered judgements the principles are supposed to account for will be solved.

Rawls presents the method as involving considerable interaction between judgements and principles. We may change our mind about a considered judgement when it does not fit with principles which account for most of our other judgements, but we may also change our mind about it when the only principles which do fit that considered judgement seem to us clearly mistaken. Hence, the set of considered judgements is not fixed as a set of data, but may change throughout the process. An important aspect of what it is for them to be judgements is that we can ask for and give reasons for them. Moreover, the person whose (moral) conception and sensibility we try to understand must be involved in this process, since only she can revise her judgements. For this person, it is a constant process of deciding what to believe. We may include in the process of RE the results of other persons' search for a reflective equilibrium, but it is still each of us that must carry on the process (Scanlon [2003] 2006: 148–9).

Now, the reasons why Scanlon thinks that any other alternative to the MRE is an apparent alternative, an illusion, are provided in his discussion of 'two commonly heard objections', which he calls the Charge of Conservatism and the Charge of Relativism (Scanlon [2003] 2006: 149). The charge of conservatism starts from the claim that, in order to decide what to believe, we should have a standard for the evaluation of the beliefs we happen to have; however, the objection goes, RE is conservative, since the principles which can be used for evaluation are formulated on the basis of our current beliefs. Whereas the MRE can be used to eliminate contradictions between beliefs, make them consistent and systematise them, the resulting principles lack the independence needed to give them real critical and justificatory force.

Scanlon's reply, on behalf of Rawls, starts with a reminder that considered judgements are not to be understood as fixed, but as open to constant revision. This would still not answer the conservatism charge satisfactorily, were it not for an additional feature of Rawls's MRE. For Rawls, considered judgements need not be simply judgements about what is just or unjust in particular cases, but are also judgements at all levels of generality, including general principles and views on the purpose of justice. Rawls talks about a Wide Reflective Equilibrium, which is formed after we examine other plausible conceptions and their supporting grounds – at the limit, it is an equilibrium which includes the conception or

set of conceptions which survives rational consideration of all feasible conceptions and the arguments in their support (Rawls 1974: 9; Scanlon 2006: 150).

Scanlon concludes:

This breadth deprives the charge of conservatism of its force. Conservative as opposed to what? It is difficult to imagine what source of criticism or justification is envisaged that the method of reflective equilibrium, so understood, would exclude. (Scanlon [2003] 2006: 150)

The suggestion here seems to be that, since everything is open for consideration and since this consideration can be undertaken from the perspective of any plausible view and argument, no source of criticism is excluded and, hence, no trace of conservatism can be found in the MRE. As Scanlon notes, however, this powerful reply leads to a different worry, which he calls the Emptiness Charge (which turns out to be a weaker version of the relativism charge): the MRE 'becomes simply the truism that we should decide what views about justice to adopt by considering the philosophical arguments for all possible views and assessing them on their merits' ([2003] 2006: 151). Yet, he thinks the emptiness charge is mistaken, since Rawls's method would be incompatible with some views about the sources of justification. For instance, it is incompatible with the view that such a source (either considered judgement or principle) could be established independently from the MRE on some other basis; and it is incompatible with the view that any class of considered judgement should be left out of the process ([2003] 2006: 151).

Consider now the more general charge of relativism: imagine two people who begin with different sets of considered judgements, but who arrive at equilibria which include different and even incompatible principles; this might suggest the principles, although incompatible, are nevertheless equally justified (which implies relativism). However, Scanlon notes that the MRE is not committed to this conclusion; given the Rawlsian distinction above between a justified principle and a person's being justified in holding a principle, we can agree that both persons are justified to hold different and even incompatible principles without asserting that the principles are equally justified. We can then examine where the difference stems from: different starting points (that is, different considered judgements)? If so, which considered judgements should be retained? Different choices at a later stage? If so, were these reasonable and should I revise

my decisions? Different sets of eligible conceptions of justice? If so, should I consider any further possibilities I might have missed?

Answers to these questions may help us to differentiate between principles as better justified and may help us revise principles and reduce the differences between them. Yet, Scanlon adds, even if I realize that both I and the other person started from reasonable considered judgements and made reasonable decisions, although the difference between the principles we support would call into question the objectivity of our moral beliefs, all we could conclude is that the MRE does not beg the question against scepticism (Scanlon [2003] 2006: 152–3).

Given that the MRE can answer these two objections and given that it excludes other justification sources which are independent from the MRE, it follows indeed that the MRE is the best and the only method of justification available. But is it really?

3. The Lack-of-Objectivity Charge to the MRE

Scanlon's presentation of the MRE and his claims that this is the best and only method of justification raise several questions. Consider, for instance, the situation at the third stage of the MRE: here the task is to reconcile the divergences between the principles formulated at stage two, in order to account for the considered judgements identified at stage one. The process of reconciliation can take the form of discarding the considered judgements which are in conflict with the principles or reformulating the principles with which considered judgements conflict or a combination of these two.

But how do we decide which of these alternatives to choose? Should I abandon the unruly considered judgements, should I reformulate the principles or a bit of both? We are told that the set of considered judgements with which we begin is open for revision. We are also told that the principles are open for reformulation, if it turns out they cannot account for what we take to be fundamental cases or if they cannot cohere with what you take to be important general principles or criteria. For instance, we (or, more exactly, those of us who have this considered judgement) can consider whether to abandon the view that distributive justice has something to do with desert or to reformulate Rawls's Difference Principle or to do a bit of both (say, abandon the considered judgement that justice is getting what one deserves and include in the Difference Principle

a note that the egalitarian starting point and the justification of differences as benefiting the worst off are ways in which we can make sure persons in the well-ordered society are not affected by undeserved factors).

Because there is no further suggestion about some general criterion or approach for such cases, the decision to commit to some conception or other appears at worst arbitrary and at best subjective. What I have just formulated here,⁷ namely, the so-called lack-of-objectivity charge, seems to be a particular instance of Scanlon's relativism charge. As we have seen, there is one weak version of this objection, the emptiness charge (that the MRE is the truism that we should decide what views about justice to adopt by considering the philosophical argument for all possible views and assessing them on their merits). Scanlon thinks this weak relativism charge is wrong, since the MRE is incompatible with some alternatives. For instance, the method is incompatible with the view that considered judgements or principles could be established independently from the MRE on some other basis and with the view that any class of considered judgements should be left out of the process. But I think there are some obvious problems with Scanlon's reply.

First, Scanlon's claim that the MRE is not vacuous, because it excludes the view that any class of considered judgements should be left out of the process, is missing the point. The claim that we should not leave out any class of considered judgements reasserts one aspect of the method Scanlon formulated as a truism above: that we should consider the philosophical argument for *all* possible views (including all classes of considered judgements) and we should be assessing them on their merit. Hence, that we cannot exclude any particular class of considered judgement does not make the MRE less vacuous, but seems to reaffirm its character as a truism.

Similarly, the other view, which Scanlon thinks is incompatible with the MRE, the view that a considered judgement or principle could be established independently from the MRE, simply reaffirms that we need to consider *all* judgements and principles when employing the MRE. If we had a source of justification independently from the MRE, then the MRE would not need to consider this source. By excluding this view, we broaden again the scope of the MRE, which can prompt again accusations that it is a vacuous method and a truism.

The fact that there is no source of justification independently from the MRE compounds the problem I mentioned above, namely, the lack-of-

objectivity charge. For, if any source of justification has to be the result of the application of the MRE, then a justification of a choice between various ways of applying the method at stage three would also need to be the result of the application of the MRE. But, then, it seems that, in order to be able to apply the MRE, we need to solve a problem which can only be solved if we can already apply the MRE.

The lack-of-objectivity charge, which I have just mentioned, is distinct from the emptiness charge discussed by Scanlon. The lack-of-objectivity charge raises a specific question concerning the third stage of the MRE – the problem is that, as formulated by Scanlon, the MRE provides us with no indication of how to decide whether to reconcile divergences between considered judgements and principles by abandoning the former, reformulating the latter or combining these alternatives. By contrast, the emptiness charge raises a general problem about the content of MRE, more exactly, whether the MRE does have any content which would be useful as part of a method of justification. Moreover, although I think the lack-of-objectivity charge is a particular case of the emptiness charge, it is a special particular case, since even if Scanlon's replies to the emptiness charge would answer it, they would not be able to answer the lack-of-objectivity charge.

The lack-of-objectivity charge makes more evident also a further problem with the MRE in relation to Scanlon's discussion of the charge of relativism. According to the charge of relativism, it would in principle be possible for two persons to start with reasonable considered judgements and to make reasonable decisions with regard to the same issue, and nevertheless end up with different principles, which are however equally justified. When these principles are incompatible, then the objectivity of our moral beliefs seems to be undermined. Scanlon does not deny this, but he takes this to be an indication of the fact that Rawls is not begging the question against the sceptic.

In other words, Scanlon's argument seems to be the following: if a sceptic questions the ability of the MRE to provide justification for objective moral beliefs, then an appropriate answer must be one which does not beg the question against the sceptic; it follows that, in principle, we should be able to conclude either that the MRE can justify objective beliefs or that it cannot. Until the sceptic's argument is proven correct (until we find the two persons who end up with equally well justified, but nevertheless incompatible, principles), all we can conclude is that the MRE does not beg the question against the sceptic

(and presumably that, for all we know, our moral beliefs have objectivity).

The fact that the MRE does not beg the question against scepticism seems to be a positive aspect of the Rawlsian method, but there are in fact two reasons why Rawls may not beg the question against the sceptic in this instance. First, it might be the case that the view that the MRE provides objectivity for our moral claims does not exclude *a priori* the possibility that our moral beliefs might not be objective. Secondly, however, it might be the case that the view that the MRE provides objectivity for our moral claims cannot be shown to be either true or false. In this case, while the way the view is formulated does not beg the question against the sceptic, neither does it have a chance to be shown to be correct or incorrect. And, yet, in *Political Liberalism*, Rawls does think he can show that his political conception of justice does have objectivity.

There are several views which Rawls considers to be undecidable and which he thinks a political conception of justice would better avoid. In fact, the distinction between a comprehensive and a political conception of justice can be drawn by reference to the issues to which the comprehensive conception is committed, while the political one avoids them and with regard to which it remains non-committal. I will first offer an example of such an issue, commitment or non-commitment to which can distinguish between a comprehensive and a political view, and then focus on the question of the objectivity of our moral beliefs.

In §5, ‘Three Conceptions of Objectivity’, of the third lecture (‘Political Constructivism’) of Rawls’s *Political Liberalism*, rational intuitionism, Kant’s moral constructivism and Rawls’s ‘political constructivism of justice as fairness’ are presented as associated with three different types of objectivity (Rawls [1993] 2005: 110). Rawls identifies five essential elements of objectivity and thinks that the three different views of objectivity are the result of the fact that rational intuitionism, Kantian constructivism and political constructivism do not understand the five essential elements of objectivity in the same way. For my purpose here, the more important elements are the first two.⁸ The first essential element of a conception of objectivity is the condition that there is

a public framework of thought sufficient for the concept of judgement to apply and for conclusions to be reached on the basis of reasons and evidence after discussion and due reflection. (Rawls [1993] 2005: 110)

Rawls thinks this framework is necessary in order for us to reach agreement by the free exercise of our powers of judgement, as opposed to a process based on mere rhetoric or persuasion. He also thinks that this public framework which enables us to reach conclusions makes possible or perhaps includes ‘mutually recognised criteria and evidence’ (Rawls [1993] 2005: 111).

The second essential element of a conception of objectivity is a corollary of the first: given that part of what we understand by a judgement is that it aims at being correct/true/reasonable (depending on the conception of objectivity we have in view), a conception of objectivity needs to include a concept of a correct judgement in accordance with its norms. For instance, for rational intuitionism Rawls notes that judgements are true by reference to an independent order of values. Political constructivism understands correct judgements as reasonable,

that is, as supported by the preponderance of reasons specified by the principles of right and justice issuing from a procedure that correctly formulates the principles of practical reason in union with appropriate conceptions of society and person. (Rawls [1993] 2005: 111)

Hence, political constructivism does not talk about moral truth, but about reasonableness. It regards a judgement as reasonable when there are many reasons in its support, reasons which are based on principles of right and justice derived (‘correctly’) from principles of practical reason (together with ‘appropriate’ conceptions of society and person).

As already mentioned, a moral theory will have a specific conception of objectivity, depending on how it understands the five essential elements which any conception of objectivity must include. For instance, for rational intuitionism, the second element specifies that a correct judgement is true in virtue of an independent order of moral values. According to Rawls, neither Kant’s moral constructivism, nor political constructivism regards moral judgements as objective in this way, since neither asserts an independent order of values. Moreover, Rawls notes, political constructivism does not deny this order of values either.

Hence, rational intuitionism is committed to an independent order of values, whereas Kantian constructivism is committed to the claim that an independent order of values does not exist. By contrast, political constructivism remains non-committal: it neither asserts this order, nor

denies it. Therefore, it is not surprising that Kantian constructivism, rational intuitionism and political constructivism have different views of objectivity. Let us now go back to the MRE and the relativism charge.

As we have seen, on Scanlon's account, the possibility of an argument which shows that our beliefs, as identified with the help of the deliberative function of the MRE, have no objectivity only indicates that Rawls does not beg the question against the sceptic. Whether actually it turns out that an argument against the objectivity of our moral beliefs can be constructed remains to be seen. But let us consider Rawls's discussion of objectivity in relation to the charge of relativism. As we have seen, one essential aspect of objectivity is an account of the correctness of a judgement, and this account of correctness is formulated in accordance with some norms. Rawls mentions that, for rational intuitionism, a correct judgement is understood as true in accordance with an independent order of values. For political constructivism, a correct judgement is defined as a reasonable judgement as supported by many reasons specified by the principles of justice derived through a procedure which correctly expresses the principles of practical reason in the context of additional appropriate conceptions of society and person. Finally, for Kant's constructivism, a correct judgement will be a right judgement as given by the categorical imperative as expression of common human reason (Rawls [1993] 2005: 115).

But what are the norms on the basis of which we can understand a judgement's correctness in the context of the MRE? As the lack-of-objectivity charge has shown, there is no indication of what is to be done, in the case of divergences between considered judgements and principles. In fact, both considered judgements and principles are open to revision, in which case, however it is decided whether to reformulate the principle or to discard the judgements which are not fitting, the decision is at worst arbitrary and at best subjective.

This seems to provide a different picture of the implication of the charge of relativism. Scanlon thinks the charge of relativism indicates that Rawls does not beg the question against the sceptic and, until the sceptic shows his argument works, we can assume our moral beliefs have objectivity. The lack-of-objectivity charge, however, suggests that we cannot account for the objectivity of our moral beliefs on the basis of the MRE and we cannot rely on this until proven wrong. Hence, while it may be true that Rawls does not beg the question against the

sceptic, it is no longer the case that we can assume objectivity until the sceptical argument is shown to succeed. It also follows that, in fact, we cannot decide in a justified way whether our moral beliefs are objective or not.

If my argument is correct, then one implication is that the MRE cannot function deliberatively – it does not enable me to commit to a particular conception in a justified way. Hence, we seem to be left with the descriptive function of the MRE. The moral theorist is an observer of moral conceptions and attitudes. These are moral conceptions and attitudes which people (including the observer herself) happen to hold; the question of their justification is suspended. Pace Scanlon, it is therefore not surprising that the later Rawls has been charged with the objection of relativism and has been interpreted as putting forward a version of pragmatism.⁹

Scanlon is right that the MRE is a method of justification in the second sense of the term (that is, descriptively); he is also right that the MRE presents a general view of a method of justification in the first sense of the term (that is, deliberatively). Yet, although the MRE does offer a good illustration of the deliberative process through which persons commit to their views, there is no indication of how the deliberative process should be concluded. As there is no additional source of justification apart from the MRE and as the MRE cannot function as a source of justification without an additional criterion, we are caught in a vicious circle.

Let us note here, therefore, that a missing element in Rawls's account of justification is a valid criterion or a set of valid criteria of justification. Rawls gets out of the circle by starting from certain fixed points, for instance, the values implicit in the public political culture of the society or the framework for the derivation of the principles of justice (which in their turn eventually are based on those values). To be sure, these 'fixed' points are in principle open for revision, as Rawls specifies explicitly, but he takes them as stepping stones for his argument. For instance, about the original position and the MRE, he says:

I shall not, of course, actually work through this process [of identifying the original position through the MRE]. Still, we may think of the interpretation of the original position that I shall present as the result of such a hypothetical course of reflection.
(Rawls [1971] 1973: 21)

Similarly, this is what he says about the idea of society as a fair system of cooperation, which he thinks is the fundamental idea of the political tradition on which his account is grounded:

the fundamental organising idea of justice as fairness, within which the other basic ideas are systematically connected, is that of a society as a fair system of cooperation over time, from one generation to the next. We start the exposition with this idea, which we take to be implicit in the public culture of a democratic society. (Rawls [1993] 2005: 15)

In both cases, an initial decision is made in order to get the process started. These starting points are not set in stone, but they are not examined or justified either. The subsequent development of the conception of justice will be conditional for its validity on the validity of the starting points, which are ultimately morally arbitrary. Again, as mentioned before, what the MRE seems, therefore, to miss is a valid criterion (or set of criteria) of moral justification.

Contrast this with a different approach in meta-ethics, the Critical Method (or CM): as part of this approach, in accordance with Scanlon's reading of the MRE, we would still have to consider which position is the most compelling in order to be able to commit to it; we would also take into consideration any relevant case, principle or theory which may help us with our evaluation of the position or positions under consideration. But the actual process of justification of the CM is distinct from that of the MRE. The MRE begins with some considered judgements and tries to formulate principles which account for the judgements. By contrast, the CM starts from the idea of an evaluative process, that is, the process of evaluating rules of action, principles or theories, and will try to determine the necessary conditions which make possible this process.

Among the necessary conditions of evaluation there must also be a criterion or set of criteria of evaluation from the perspective of which particular positions can be questioned, challenged and either rejected or vindicated. This is precisely the criterion or set of criteria missing from Rawls's account. The approach used by the CM is sometimes called constitutivist, sometimes transcendental, and it is usually traced back to Kant. In the next section, I would like to examine the prospects of supplementing Rawls's political constructivism with some Kantian elements. This should give us a better perspective from which to answer

the lack-of-objectivity charge (together with the relativism and conservatism ones), as well as a more promising basis for an account of moral justification.

4. Universal Principle of Right

In Kant's political philosophy, the fundamental criterion for the justification of principles of law is the Universal Principle of Right:

Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law. (*MM*, 6: 230)¹⁰

This seems to provide a criterion which Kant claims to be able to justify by following the CM.¹¹ One immediate worry we may have when trying to supplement Rawls's political constructivism with Kant's Universal Principle of Right¹² is that, as already mentioned, a decisive reason for the political turn Rawls takes in his later writings is given by the desire to distinguish his account from Kant's and, in particular, to distinguish between Kant's comprehensive intentions and his own, political purpose.

According to Rawls, there are three main differences between what he calls 'Kant's moral constructivism' and his own 'political constructivism of justice as fairness' ([1993] 2005: 99).¹³ The first difference is given by Kant's view of autonomy, which Rawls thinks has in Kant a regulative role for all of life ([1993] 2005: 99). A second difference is given by the ground of Kant's moral philosophy, transcendental idealism ([1993] 2005: 99–100). Finally, there is a difference in the aims with which the two accounts proceed ([1993] 2005: 100–1). According to Rawls, any of these differences is 'far-reaching enough to distinguish justice as fairness from Kant's moral constructivism' (Rawls [1993] 2005: 101).

In what follows, I aim to show that the first and third differences are not really as far-reaching as Rawls suggests, and that the second one, while far-reaching enough, nevertheless, is needed as part of Rawls's political constructivism, if he wants to be able to provide an account of objectivity; this, however, will not affect political constructivism's neutrality – in fact, there can be no argument to the neutrality of political constructivism without an account of the objectivity of our moral beliefs.

It seems quite clear why Rawls abandons the Kantian ground of his earlier *A Theory of Justice* (1971), in order to focus on the political aims of his constructivism in *Political Liberalism*. As we have seen, the Kantian account includes views of the aim of philosophy, the metaphysical ground of moral theory and the fundamental value of human life, which are more comprehensive than Rawls thinks is needed for political purposes. Given the controversial nature of some of Kant's claims (for instance, his transcendental idealism), by including the Kantian background as a condition of political constructivism, the support that political constructivism would be likely to receive from the members of a well-ordered society would be significantly reduced. Moreover, Rawls's theory of justice would seem to support a particular group of conceptions of the good, which are compatible with the Kantian framework, while disadvantaging many of the other groups with conceptions of the good which are incompatible with Kantianism. In this way, the neutrality of Rawls's theory of justice would also be undermined.

However, consider first Kant's idea of autonomy. Although this is a notion which seems to play a fundamental role in Kant's ethics, particularly in the *Groundwork* and in the *Second Critique*, it gradually disappears in the 1790s from Kant's writings, for instance, from the 'Doctrine of Virtue'. This seems to undermine claims to the fundamental role of this notion in Kant's philosophy in general, and in his ethics in particular.¹⁴ If Kant's ethics survives largely unchanged without the notion of autonomy, then this notion cannot be taken to be so central for Kant's account and, hence, Rawls should not take it as a distinguishing element of Kant's thought. Let us suppose, however, that we focus on the *Groundwork* and *Critique of Practical Reason*, where autonomy does seem to play an important role.

The first documented public reference to the ideal of autonomy occurs in the Introduction to the *Feyerabend Lectures on Natural Law* (1784), which are lectures in political theory. There are similarities between Kant's account of ethical autonomy and his account of political legislation. For instance, in the same way in which a political legislator gives law to the entire people, rather than primarily to himself, the ethical agent gives law to the entire community. To be sure, the resulting law will also apply to the legislator. Hence, autonomy means being subject to laws that are one's own, not laws that are given by another.

Hence, the Formula of Autonomy enjoins agents to regard themselves as giving universal laws, to which they themselves are subject. This does not

actually mean formulating some new ethical laws but testing maxims for their permissibility. This shows that Kant's idea of autonomy does not have voluntarist implications – the content and validity of the ethical laws do not depend on an act of the agent or on her actual consent. Moreover, the agent cannot rescind ethical laws either. The will of the agent has the property of autonomy, in the sense of its being subject to its own laws, but the obligatory force of moral duties does not derive from a contingent act of the agent.¹⁵

There are two significant aspects which need to be emphasized here. First, consider the idea of autonomy as being subject to one's own laws, but not in a voluntarist sense: the non-voluntaristic account claims that the idea of regarding myself as creator of my own laws refers to the fact that I can acknowledge the validity of the law and, hence, that I am subject to the law with the appropriate (ethical) motivation. In other words, I act in accordance with and for the sake of that law. This represents nothing other than the famous Kantian requirement concerning motivation and moral worth in ethics.¹⁶ Secondly, in addition to the condition of ethical motivation, the idea of autonomy only imposes a condition of universality.

These two aspects show that, even in the case of Kant's ethics in the *Groundwork* and the *Second Critique*, the idea of autonomy does not add to Kant's philosophy features which would considerably distinguish Kant's political position from Rawls's political constructivism. This is because, first, as it is well known, in Kant's political theory, unlike his ethical theory, there is no condition imposed on motivation. Secondly, the universality which is required by autonomy is also required as part of Kant's political theory, but the same can be claimed about political constructivism. This is clear from Rawls's discussion of objectivity: the idea of universality both in Kant and Rawls is not the idea of a unique set of laws, which are valid for all societies; it is the idea of a set of laws, which is just for a particular society, given its specific features and circumstances.

Consider now the third aspect of Kant's philosophy, which Rawls thinks distinguishes sharply Kant's political theory and Rawls's political constructivism. According to Rawls, Kant takes the aim of philosophy to be

that of showing the coherence and unity of reason, both theoretical and practical, with itself; and of how we are to view reason as the final court of appeal, as alone competent to settle all questions about the scope and limits of its own authority. . . .

Kant shifts the burden of proof: the affirmation of reason is rooted in the thought and practice of ordinary (sound) human reason from which philosophical reflection must begin. Until that thought and practice appears to be at odds with itself, it needs no defense. (Rawls [1993] 2005: 101)

Several things are worth noting here. First, as Rawls acknowledges, it is difficult to present briefly what Kant takes the aim of philosophy to be. Kant's philosophy has a systematic character and does not limit itself to particular areas (say, political philosophy or ethics or epistemology); hence, any overall characterization of the aim of the Kantian project will be a complex enterprise. Taking Rawls's brief presentation of the aims of Kant's philosophy as decisively in contrast with the aims of Rawls's political constructivism is, on the face of it, unwarranted.

Secondly, at least part of Rawls's description of Kant's view of the aim of philosophy seems to be challenged by alternative readings. For instance, consider the suggestion that Kant takes reason to be the final court of appeal and alone competent to settle all questions about the scope and limits of its own authority. This is reminiscent of a 'traditionalist' reading of Kant, which seems to assume 'the underlying framework of a normative, antiskeptical, foundationalist epistemology' (Bird 2006: 9). By contrast, a 'revolutionary' reading takes Kant to anticipate a Wittgensteinian 'therapeutic aim', namely, 'to diagnose and correct errors in the philosophical tradition' (Bird 2006: 15) and to offer 'a descriptive survey and classification of the central, fundamental elements constituting our experience' (Bird 2006: 10).

Insofar as the latter reading portrays Kant's political philosophy as closer to Rawls than the former does, we have again an example of the potential compatibility, rather than of the radical difference, between their views of the aims of philosophy. Moreover, this is complemented by Rawls's emphasis, in his presentation of Kant's view of the aims of philosophy, on ordinary human reason, an aspect which seems more akin to Rawls's political constructivism rather than to a Kantian system of philosophy.

Thirdly, the difference between Rawls and Kant is not placed, in Rawls's account, on the proper level. There is no question that the scope of Rawls's argument is intentionally much narrower than that of Kant's philosophy. Hence, what we would need for an appropriate comparison between Kant and Rawls is a focus on the aims of Rawls's political constructivism and of Kant's theory of justice, rather than the aims of

Rawls's theory of justice and of Kant's philosophy as a whole. As we have seen, Rawls states that the aim of his theory of justice is to defend, 'given reasonably favourable conditions', 'the possibility of a just constitutional democratic regime' (Rawls [1993] 2005: 101). I think Kant would certainly agree with this, as one of the important aims of his political philosophy (although he would probably label this regime 'republican', rather than 'democratic').

Fourthly, in spite of the contrast that Rawls's presentation of the aims of Kant's philosophy is meant to create with political constructivism, there are obvious similarities, which it is difficult not to notice: the coherence and unity which Kant seems to seek for reason are very similar to the coherence and unity which the MRE is supposed to bring about between considered judgements; the emphasis on ordinary sound human reason, as I have mentioned, resembles Rawls's focus on those common aspects of our political culture which make possible an overlapping consensus; and the shift to an assumption that the thought and practice of ordinary human reason need no defence is similar to Rawls's taking the values of our public political culture as starting points.

All these considerations show that the project of a political constructivism enriched with the Kantian element of the Universal Principle of Right or of a similar principle performing the same function is less likely to be undermined by a stark difference between the views to be combined. There is, however, one final aspect of Kant's philosophy which seems indeed too foreign to Rawls's philosophy to allow for such a Kantian political constructivism – this is the second aspect mentioned above, Kant's transcendental idealism. Rawls is right that Kant's moral philosophy is grounded in transcendental idealism (or at least this is what I assume for the purpose of this article). This is what makes it possible for Kantian moral constructivism, unlike Rawls's political constructivism, to express a commitment to the non-existence of the rational intuitionist's independent order of values. Moreover, this is what makes it possible also for Kantian moral constructivism to provide an account of moral objectivity centred on the Universal Principle of Right.¹⁷

Hence, adding the Universal Principle of Right (or a similar principle) to Rawls's political constructivism is going to commit a Rawlsian metaphysically to more than she intended. Yet, given Rawls's intention to provide an account of moral objectivity and the failure of political constructivism to offer such an account, it would be important at least to relax the strategy of avoidance to allow some hypothetical philosophical engagement with

some of the positions which are currently avoided by Rawls's political constructivism as comprehensive, and it would be significant to allow some openness to a possible more substantial metaphysical commitment, when it is the result of a critical discussion of the positions available.

The argument so far shows that political constructivism can be open to consideration of some of the positions in philosophy and meta-ethics which it has so far tried to avoid. This is made necessary by the internal inconsistency in the later Rawls between a commitment to the objectivity of moral justification and the failure of the MRE to provide objective justification. If correct, my argument shows that some of the differences which Rawls considers as crucial between his political constructivism and Kant's philosophy, particularly, his philosophy of right, are much less stark than it may seem as the beginning. From this perspective, the idea of a Kantian political constructivism which adopts Kant's Universal Principle of Right or a similar principle seems a feasible project.

Yet, as we have seen, the supporter of the Rawlsian theory of justice might, as a result, have to accept a commitment to some metaphysical claims, which she initially might have liked to avoid. If such commitments are the result of a critical consideration of some of the positions in meta-ethics and philosophy which can account for moral objectivity, then they will not affect the Kantian political constructivism's neutrality: in fact, without an account of the objectivity of moral justification, no argument to the neutrality of political constructivism would be possible.

5. Conclusion

In his attempt to reconcile conflicts between the conceptions of the good of members of a well-ordered society, Rawls expresses a commitment to the objectivity of the judgements and principles with the help of which such conflicts can be adjudicated upon. There is here a commitment to justice, understood as provided by an objectively justified framework of principles and laws, as a condition of reconciliation and peaceful coexistence. Yet, as we have seen, Rawls's account of justification and in particular the central method of this account, the MRE, are unable to provide the conception of objectivity that Rawls thinks appropriate for his political constructivism.

Literature on the MRE, as a general justificatory strategy in meta-ethics, points to the inability of the MRE to provide an account of objective

moral justification and, in some cases, suggests as an alternative the transcendental method, particularly the constitutivist version. In this article, I have examined the extent to which Rawls's MRE is able to account for objective justification within the context of his political liberalism, and in particular to account for the specific type of moral objectivity he thinks apposite for his political constructivism. I have shown that, even on the basis of the strongest interpretation and defence of the MRE, Rawls is left with an internal inconsistency between the commitment to objectivity and the inability to provide an account of objective moral justification.

I have suggested that one alternative would be to adopt Kant's Universal Principle of Right as an objective criterion for the justification of principles of law and of specific laws. I have then examined the extent to which Rawls's shift, in the later writings, to a political liberalism which avoids a commitment to Kant's philosophy does not make impossible the project of a Kantian political constructivism centred on the Universal Principle of Right or on a similar Kantian principle.

I have argued that the differences between political constructivism and Kant's philosophy, which Rawls presents in *Political Liberalism*, are much less stark than they may initially seem. The only important difference which might undermine this project is Kant's commitment to transcendental idealism. But I have claimed that, insofar as this is the result of a critical examination of some of the positions concerning objectivity in meta-ethics and philosophy more generally, it will not affect Rawls's commitment to neutrality, but, on the contrary, will make such a commitment possible.

I conclude with some clarification of the new model of justification I propose, in response to some possible objections.¹⁸ There seems first to be some confusion with regard to the status of the UPR in relation to the MRE: is it supposed to be a fixed feature of a revised MRE process or should it be a criterion or principle practitioners of the MRE should be entitled to entertain, rather than simply dismiss as comprehensive? In response to this, I first note that it is not a matter of an either/or: the claim is that the UPR can be justified by an alternative method, the CM; if this is correct, then, once practitioners of the MRE are allowed to consider it, given the problem raised by the lack-of-objectivity charge, they will get to accept the UPR as justified and, hence, as a fixed feature, which they can use in making decisions at the third stage of the MRE.

This is a fixed feature not in the sense that it cannot be challenged – as I have just said, we take the UPR to be right, because it is justified through

the CM; hence, it can always be challenged, if its justification seems unconvincing; the UPR is, however, fixed, in the sense that it cannot be challenged by considered judgements or by those principles which are formulated with the help of Rawls's MRE to account for considered judgements. This is because, as we have seen, the starting point for the CM is not a particular set of considered judgements and the corresponding principles, but the general idea of evaluation (of assessing the validity of rules or actions, principles or theories), in particular in legal and political philosophy. In theory at least, in order for a considered judgement or a principle to pose a challenge, they will need to presuppose the criterion from the perspective of which the critical and justificatory process is possible, and this is precisely the UPR.

A related question concerns the status of the UPR as far as the level at which it is supposed to work is concerned. Thus, I have suggested that the UPR is a criterion for the evaluation of rules of action, principles and theories. But would this not lead to an infinite regress? If the dissonance between principles and judgements needs a criterion of adjudication, would a further norm not be needed to adjudicate between potential discrepancies between principles/judgements, on the one hand, and, on the other, the criterion (UPR)?

First, conflicts between principles/judgements and the UPR are not possible, if what I have just said about the CM and the UPR is correct. Secondly, however, even if such a conflict were possible, there would not be a higher-order criterion on the basis of which the conflict would be solved; as a necessary condition of evaluation, the UPR is the most general criterion of justification in legal and political philosophy, so any challenge to this principle would need to be considered by examining whether it is indeed a necessary condition of evaluation.

A second objection grants that we can perhaps reconcile a view of the UPR as a relatively fixed criterion of justification with a conception of the UPR as an element which practitioners of the MRE are entitled to entertain; if the UPR is a new feature of a revised MRE, a feature which cannot be discarded, then it is unclear what such a revised MRE would have as distinctively Rawlsian about it. This question is particularly difficult, since it involves the use of labels such as 'Rawlsian' and 'Kantian' which cannot be easily separated. How would an account of moral justification need to be in order to qualify as distinctively Rawlsian? The context of the question is given by the revised MRE, which incorporates some Kantian elements. Hence, the question asks us to identify elements

which are Rawlsian and not Kantian. But Rawls is generally viewed as a Kantian philosopher.

It follows that the focus will have to be again on some of the three elements which Rawls thinks distinguish his account from Kant's, as well as on the concerns from which those three elements derive. Recall the three elements: autonomy, transcendental idealism and the aim of philosophy; all these elements derive from an acknowledgement of, and concern for the implications from, the fact of pluralism.

We have already seen how this concern for pluralism, which we can take to be a distinctively Rawlsian element, can help us support a different reading of the role of autonomy in Kant, as well as of his view of the aim of philosophy. Moreover, we have seen how it leads to an understanding of transcendental idealism as a metaphysical position which is the result of an engagement with alternative philosophical positions on objectivity. The Rawlsian MRE is not replaced by a distinct Kantian framework; as we have seen, the Rawlsian MRE is supplemented with an account of a criterion of justification (the UPR), which is derived through a distinct method (the CM) and which enables Rawls to offer an account of the objectivity of moral beliefs within the constraints imposed by his concern for pluralism. The resulting revised MRE is still neutral towards the various conceptions of the good, even if it is more committed philosophically.

A third and final important objection to be mentioned here grants both that the revised MRE includes a relatively fixed element, which practitioners of the MRE should be allowed to entertain, and that the resulting revised MRE preserves some distinctively Rawlsian elements; what it questions is the extent to which the UPR is consistent with Rawls's own principles of justice, where consistency here is considered along the lines of a logic of moral justification in Kant. Thus, according to Kant, the UPR should be limited to the form of principles of justice and should not include any specific content pertaining to such principles. Yet, Rawls's Difference Principle and perhaps even the Principle of Fair Equality of Opportunity seem to violate precisely this condition.

To answer this, I only need to make explicit some aspects of the revised MRE, which I have already introduced in this article. Thus, we have seen that the Kantian criterion of justification which is needed for an appropriate account of the objectivity of our moral beliefs *complements* Rawls's MRE. The UPR and its justification are based on a different method of justification, the CM, which supports a criterion of

justification able to adjudicate in cases of conflicts between principles of justice and considered judgements. Hence, the criterion is of a higher order than principles and considered judgements, and is not inconsistent with the Rawlsian principles of justice.

The Rawlsian principles of justice are formulated for a particular type of society and given a particular public political culture. This is compatible with a view of the Rawlsian MRE as a strategy for the formulation of principles of justice for specific types of society, and compatible with a view of the UPR as a criterion for the adjudication of conflicts which arise when principles of justice formulated with the help of the MRE conflict with considered judgements.

In short and going back to the distinction between two notions of justification introduced at the beginning of this article, the MRE concerns the second, descriptive notion of justification; by contrast, the CM refers to the first, deliberative notion of justification. The revised MRE includes the Kantian CM and the Rawlsian MRE; it preserves distinctively Rawlsian elements and, when entertained by practitioners of the MRE, it can be shown to be justified on the basis of the CM.

Notes

- 1 This article was written during an institutional research leave ('sabbatical'), when I was a Visiting Professor in the Department of Philosophy at the University of Warwick, a Visiting Researcher at the Oxford University's Uehiro Centre for Practical Ethics, and holder of a British Academy Newton Advanced Fellowship. I am grateful to my home university, to the two host institutions, and to the British Academy for making this period of research leave possible and productive. I am very grateful to the guest-editors of this special issue for very helpful comments on drafts of this article and for their patience and support throughout the editorial process.
- 2 See for instance, Hare (1973), Singer (1974) and Brandt (1979).
- 3 See, for instance, Daniels (1979), Brink (1989), Scanlon ([2003] 2006) and Walden (2013).
- 4 See, for instance, De Maagt (2017); although De Maagt's aim is not to reject altogether the MRE, he does suggest an alternative method of moral justification, namely, that which is at the basis of 'transcendental arguments' (2017: 445). He refers mainly to arguments usually called 'constitutivist'.
- 5 I have mentioned (see n. 4 above) that one of the alternatives on offer is given by the transcendental strategy of constitutivist arguments. In two recent papers, I have examined the ability of two types of constitutivist argument to justify normative constraints. My answer was positive, pending some further work from supporters of constitutivism – see Baiausu (2016a and forthcoming).
- 6 For instance, De Maagt focuses on T. M. Scanlon's version of the MRE in Scanlon (2014). Another aspect which is not usually considered is given by the differences between the ways Rawls understands the MRE in his various texts. For a discussion of

- this aspect, see Mikhail (2011). In what follows, I will not discuss these differences, since my focus is on the philosophy of the later Rawls and in particular his *Political Liberalism*. ([1993] 2005).
- 7 I have already formulated this objection to Rawls's MRE in Baiasu (2001). Here I am developing that critique in the context of Scanlon's defence of Rawls's MRE and I am drawing some of its implications for the issue of justification in the later Rawls.
 - 8 In fact, as we will see, the more relevant element is the second, but I mention both of them, since Rawls takes the second element to be a corollary of the first.
 - 9 For instance, by Rorty ([1988] 1990).
 - 10 In citing Kant's works the following abbreviations are used: MM: *The Metaphysics of Morals* (*Die Metaphysik der Sitten*), comprising the *Metaphysical First Principles of the Doctrine of Right* (*Metaphysische Anfangsgründe der Rechtslehre*) (1797) and the *Metaphysical First Principles of the Doctrine of Virtue* (*Metaphysische Anfangsgründe der Tugendlehre*) (1797), in Kant (1996: 353–603); G: *Groundwork of the Metaphysics of Morals* (*Grundlegung zur Metaphysik der Sitten*) (1785), in Kant (1996: 41–108). Pagination references in the text and footnotes are to the volume and page number in the German edn of Kant's works, *Kants gesammelte Schriften* (1900–). I am using the translations listed in the References.
 - 11 As we will see below (n. 17), there is a debate in the literature on whether the Universal Principle of Right is derived from the Categorical Imperative. The interesting positions are those which claim that the Universal Principle of Right cannot be derived from the Categorical Imperative (for instance, Willaschek 1997 or Wood 2002) or that the Universal Principle of Right might be derivable from the Categorical Imperative, but can also be derived independently (for instance, Ripstein 2009). This would open up the possibility that the Universal Principle of Right would not be justified through the CM, although we know that it would need to be another strategy than the MRE. I believe that there is no such alternative, but showing this goes beyond the scope of this article. Here, I assume that the Universal Principle of Right can be derived with the help of the CM. My own view is that this happens indirectly – through the Moral Law and the Categorical Imperative.
 - 12 It might be claimed that Rawls already has a criterion of justification similar to the Universal Principle of Right, namely, his idea of justice as fairness. Although it might be the case that the contents of these two criteria of justification are equivalent and that they are meant to perform the same function in Kant and Rawls, nevertheless, as we have seen, in Rawls we cannot see this criterion as a valid objective criterion, since we cannot account for objectivity. The idea of complementing Rawls's political constructivism with the Universal Principle of Right is primarily one of adding a *valid* objective criterion, rather than that of adding to the *content* of an (unjustified) criterion already existing as part of political constructivism.
 - 13 In fact, Rawls says that there are four differences, but I only focus on those derived from features of Kant's philosophy (as identified by Rawls).
 - 14 There are only a few texts which discuss possible reasons for this change in Kant's philosophy. Pauline Kleingeld suggests that this shift in Kant's view of autonomy occurs when Kant changes his view of political autonomy and legitimacy, since she understands Kant's notion of autonomy in ethics as formulated on the basis of an analogy with political legislation (Kleingeld 2017, 2018).
 - 15 I have followed here Kleingeld's recent account of autonomy in Kant. Although I disagree with this account in some respects, for the purpose of this article I need not discuss this further.
 - 16 For instance, see G, 4: 431.

- 17 There is an ongoing debate on the relation, in Kant, between the Universal Principle of Right and the Categorical Imperative. I have defended a complex account of the relation, against views which affirm a simple relation of dependence or independence. On my account, the normative ground of Kant's account of moral justification is given by the Moral Law, from which the Categorical Imperative and the Universal Principle of Right can be derived (Baiaus 2016b, 2016c).
- 18 I am grateful to an anonymous reviewer for this journal, who formulated the following three objections.

References

- Baiaus, S. (2016a) 'Constitutivism and Transcendental Practical Philosophy: How to Pull the Rabbit out of the Hat'. *Philosophia*, 44, 1185–208.
- (2016b) 'Right's Complex Relation to Ethics in Kant'. *Kant-Studien*, 107, 2–33.
- (2016c) 'Ethical and Politico-Juridical Norms in the *Tugendlehre*'. *Studi Kantiani*, 29, 59–75.
- (forthcoming) 'Philosophical Timidity: The Constitutivist's and Naturalist's Hesitant Quest for Normativity'. In A. Pinheiro Walla and R. Demiray (eds), *Reason, Normativity and Law: New Essays in Kantian Philosophy* (Cardiff: Wales University Press).
- Bird, G. (2006) *The Revolutionary Kant: A Commentary on the Critique of Pure Reason*. Chicago and La Salle, IL: Open Court.
- Brandt, R. (1979) *A Theory of the Good and the Right*. Oxford: Oxford University Press.
- Brink, D. O. (1989) *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Daniels, N. (1979) 'Wide Reflective Equilibrium and Theory Acceptance in Ethics'. *Journal of Philosophy*, 76, 256–82.
- De Maagt, S. (2017) 'Reflective Equilibrium and Moral Objectivity'. *Inquiry*, 60, 443–65.
- Hare, R. M. (1973) 'A Theory of Justice'. *Philosophical Quarterly*, 23, 144–55.
- Kant, I. (1900–) *Gesammelte Schriften*. Ed. by the Königlich Preußischen Akademie der Wissenschaften, subsequently Deutsche, now Berlin-Brandenburg Akademie der Wissenschaften (originally under the editorship of Wilhelm Dilthey). Berlin: Georg Reimer, subsequently Walter de Gruyter.
- (1996) *Practical Philosophy*. Tr. and ed. M. J. Gregor. Cambridge: Cambridge University Press.
- Kleingeld, P. (2017) 'The Principle of Autonomy in Kant's Moral Philosophy: Its Rise and Fall'. In E. Watkins (ed.), *Kant on Persons and Agency* (Cambridge: Cambridge University Press), pp. 61–80.
- (2018) 'Moral Autonomy as Political Analogy: Self-Legislation in Kant's *Groundwork* and the *Feyerabend Lectures on Natural Law*'. In S. Bacin and O. Sensen (eds), *The Emergence of Autonomy in Kant's Moral Theory* (Cambridge: Cambridge University Press), pp. 158–75.
- Mikhail, J. (2011) 'Rawls's Concept of Reflective Equilibrium and its Original Function in A *Theory of Justice*'. *Washington University Jurisprudence Review*, 3, 1–30.
- Rawls, J. ([1971] 1973) *A Theory of Justice*. Oxford: Oxford University Press.
- (1974) 'The Independence of Moral Theory'. *Proceedings and Addresses of the American Philosophical Association*, 48, 5–22.
- Rawls, J. ([1993] 2005) *Political Liberalism*. Expanded edn. New York: Columbia University Press.

- Ripstein, A. (2009) *Force and Freedom. Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press.
- Rorty, R. ([1988] 1990) 'The Priority of Democracy to Philosophy'. In A. R. Malachowski (ed.), *Reading Rorty* (Oxford: Blackwell), pp. 381–402.
- Scanlon, T. M. ([2003] 2006) 'Rawls on Justification'. In S. Freeman (ed.), *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press), pp. 139–67.
- (2014) *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Singer, P. (1974) 'Sidgwick and Reflective Equilibrium'. *Monist*, 58, 490–517.
- Walden, K. (2013) 'In Defence of Reflective Equilibrium'. *Philosophical Studies*, 166, 243–56.
- Willaschek, M. (1997) 'Why the "Doctrine of Right" does Not Belong in the "Metaphysics of Morals": On Some Basic Distinctions in Kant's Moral Philosophy'. *Jahrbuch für Recht und Ethik*, 5, 205–7.
- Wood, A. (2002) 'The Final Form of Kant's Practical Philosophy'. In M. Timmons (ed.), *Kant's Metaphysics of Morals: Interpretative Essays* (Oxford: Oxford University Press), pp. 1–22.