Rawls' Methodological Blueprint

This is a pre-proofs, pre-copy-edited version. Please only cite the finished version, which is now published in the *European Journal of Political Theory*, and available at:
http://ept.sagepub.com/content/early/2015/09/23/1474885115605260.abstract?rss=1

Abstract

Rawls' primary legacy is not that he standardised a particular view of justice, but rather that he standardised a particular method of arguing about it: justification via reflective equilibrium. Yet this method, despite such standardisation, is often misunderstood in at least four ways. First, we miss its continuity across his various works. Second, we miss the way in which it unifies other justificatory ideas, such as the 'original position' and an 'overlapping consensus'. Third, we miss its fundamentally empirical character, given that it turns facts about the thoughts in our head into principles for the regulation of our political existence. Fourth, we miss some of the implications of that empiricism, including its tension with moral realism, relativism, and conservatism.

Catherine Audard John Rawls, *Stocksfield:Acumen*, 2007.

Samuel Freeman Rawls, *London:Routledge*, 2007.

Paul Graham Rawls, *Oxford:Oneworld*, 2007.

Percy B. Lehning John Rawls:An Introduction, *Cambridge: Cambridge University Press*, 2009.

Jon Mandle & David A. Reidy (eds) A Companion to Rawls, *Chichester: WileyBlackwell*, 2014.

Sebastiano Maffetone Rawls:An Introduction, *Cambridge: Polity*, 2010.

Thomas Pogge John Rawls:His Life and Theory of Justice, *Oxford: Oxford University Press*, 2007.

Introduction

Rawls' primary legacy is not his theory of justice, or even the debate he started about that concept, but rather the way in which that debate is carried out. In other words, his primary legacy is *methodological*, in the sense that he standardised a particular way of doing political philosophy, or, more precisely, a particular method for the justification of political principles. This is why it makes sense to say that he created a 'common disciplinary discourse' (Lehning, 2009, 14), that he gave us 'new ways to argue' (Freeman, 2007, 459), that he inaugurated a 'formidably productive paradigm in moral, social, and political philosophy' (Maffetone, 2010, 14), and that, even if his conception of justice was not particularly original, the way he argued about it was (Audard, 2007, 7; Freeman, 2007, x; Graham, 2007, 167; Maffetone, 2010, 10). Yet there is still confusion about that method, despite such standardisation. What, for example, is the precise relationship between 'wide reflective equilibrium', the 'original position', and the idea of an 'overlapping consensus'? How, if at all, does the method change between *A Theory of Justice* and *Political Liberalism*? Which, if any, of the following c-words accurately summarises its character: contractarian, contractualist, constructivist, contextualist, or coherentist?

Given such confusion, we can be grateful for the gradual emergence of a range of books on Rawls' works *when taken as a whole*, and in particular the light this approach sheds on the

continuous method underpinning those works: justification via reflective equilibrium. This, in short, is the idea that runs from the first of Rawls' works to the last, that explains, more than anything else, their fundamental character, and that became standard for a subject that was revitalised as a direct result of the promise it held. In what follows, I use these books to paint a distinctive and four-part picture of this method. First, I explain its *context*, *core*, and *continuity*. Second, I discuss its surprising *empiricism*. Third, I explore the *problems* it encounters, partly in virtue of such empiricism. Fourth, I comment on its *future*.

The context, core, and continuity of Rawls' method

We start then with the context of this method, about which one needs to understand at least two things: First, the argument of *A Theory of Justice* was developed over the 1950s and 1960s, as captured by a series of articles leading up to that book; second, this was a time when people were asking whether political philosophy still existed (Berlin, 1962), or, more bluntly, whether it was dead (Laslett, 1956). Laslett's account is especially important, given his third reason for the problem:

> 'The Logical Positivists did it. It was Russell and Wittgenstein, Ayer and Ryle who convinced the philosophers [to] withdraw […] and re-examine their logical and linguistic apparatus. [This] called into question the logical status of all ethical statements, and […] threatened to reduce the traditional ethical systems to […] nonsense. Since political philosophy is, or was, an extension of ethics, the question has been raised whether [it] is possible at all' (Laslett, 1956, ix)[1]

So, when understanding Rawls' achievement, the point is that this was a context in which it would have seemed futile, to many, to compare the relative merits of utilitarianism and a re-worked social-contract theory, given that their differences, once logical inconsistencies are removed, must amount to either disagreements of fact (and thus the preserve of science) or boo/hurrah opinion (and thus irrelevant to philosophy). On this view, there is no proper argument to be had regarding what one *should* think about the matter, or, if this is a separate idea, what a *reasonable* person *would* think about it. Instead, all that is left to moral philosophy are meta-ethical questions about the meaning of moral utterances – as illustrated by Ayer's emotivism - whilst all that is left to political philosophy – as illustrated by Berlin, Ayer's friend and colleague - is a choice between historical interpretations or a modest form of conceptual analysis, in which one can draw out the implications of different concepts, as well as the range of possible trade-offs between them, without ever making principled recommendations for how such things *ought* to be done[2].

This context of apparently restricted normativity can be described in various ways (cf. Barry, 1990; Wolff, 2011; and Floyd, forthcoming-a). One might talk of the heyday of 'ordinary language philosophy' or 'linguistic philosophy', or of 'analytic', as opposed to 'analytical' philosophy, or of course of 'logical positivism', as Laslett does. One might talk more generally of the influence of 'positivism', or more narrowly of doctrines such as 'expressivism' or 'emotivism'. Or one might talk of the dominance of normative theory by 'meta-ethics', and not just because the meta-ethical thinking of the day, as described, seemed to render meaningful argument about political principles impossible, but also because Rawls himself strove to show how the latter could be made 'independent' of the former, regardless of the positions that subject ends up adopting (Rawls, 1999, 286-302). The key thing though is this: what Rawls was doing in *Theory* was providing us with a method for meaningfully comparing the merits of different moral and political principles at a time when such comparison seemed impossible (Audard, 2007, 5; Freeman, 2007, 12; Graham, 2007, 4-5; Lehning, 2009, 12-15).

How could he do that? Clearly not by developing 'a substantive theory of justice founded solely on truths of logic and definition', for he accepted that this was 'obviously impossible' (Rawls, 1971, 51). Instead, he believed roughly the following: (1) even if our moral judgements, e.g. 'murdering Paul would be wrong', boil down to brute feelings, those feelings can still be systematised into principles, e.g. 'do not murder'; (2) even if those principles, in virtue of such feelings, look like nothing more than opinions, the opinions of different individuals can still be aligned with each other on political fundamentals – we might, after all, come to agree on 'not murdering'; and (3) even if these feelings are just 'facts' about us, this process of systematisation and alignment still counts as justification of the principles they lead to, much as laws of nature are grounded by empirical observations – though it is not an identical process, given the revisability of these facts, and the lack of any commitment to the 'truth' of such principles, as we will see below. Nevertheless, this is why it makes sense for him to call his project a 'theory' of justice, and even on occasion a 'theory of the moral sentiments' (Rawls, 1971, 51, 120).

What though, precisely, is the 'method' to which these claims lead? Is it contractarianism? (Graham, 2007, 9) Contractualism? (Pogge, 2007, 131) Contextualism? (Audard, 2007, 18) Constructivism? (Laden, 2014, 60) A better answer, on balance, is *coherentism*, given that this term highlights in Rawls' work the fundamental importance of our existing thoughts to justification, in the sense that a principle is more or less justified (than another principle) according to the extent to which it *coheres* with our current normative convictions. Alternative justificatory ideas are thus subordinate to this fundamental one, given that, for example, the *construction* of any *contract*-like situation is itself dependent on the described coherence, whilst the many positions taken on *contextualism* are simply a set of positions on whose thoughts (and what kind of thoughts) count[3].

Coherentism, however, is not the label Rawls ran with in his own work (or even one of his index entries). Instead, he articulated this general idea via the more precise notion of justification via reflective equilibrium which, as we will see, is essentially a particular method for turning a certain class of our thoughts (considered judgements) into a certain class of principles (political principles - though of course Rawls focuses on principles of justice)[4]. It was with this method that Rawls ushered in a new 'thoughts-to-oughts' era of justification for political philosophy, in which various thinkers (Nozick, Dworkin, Cohen, etc.) would put various thoughts together (intuitions, hypothetical choices, considered judgements, etc.) in order to elicit principles of varying subject matter (legitimacy, justice, rights, etc.)[5]. And it was on this method that his own work was centred, just as contemporary scholarship is centred upon it. It is, we should see, both the core of his project and the idea on which he was working from the very start of his career (Rawls, 1999, 1-19; Lehning, 2009, 5; Reidy, 2014, 12-23). Or, as Freeman puts it, justification via reflective equilibrium is not just 'the most general idea of justification', but also 'the framework for understanding those other ideas' (2007, 29).

Yet we need to be careful with it. We are ham-fisted when we simply invoke 'reflective equilibrium', sometimes as a process, sometimes as the end-point of a process, and without any distinction between its different types – narrow/wide/general/full (to which we are about to come). Instead, we understand it properly only when we understand its five distinct stages[6], which run as follows. First, you gather up your 'considered judgements' - those convictions of justice/injustice of which you're most confident (though if your subject was something other than justice, you would start with a different set of judgements). The

paradigmatic case is the injustice of slavery – Rawls liked Abraham Lincoln's claim that if slavery is not wrong then nothing is wrong (Lehning, 2009, 1; Maffetone, 2010, 2), and it is worth noting that all his key examples were of *in*justice rather than justice, e.g. serfdom, religious persecution, and the oppression of women (Rawls, 2005, 431; Audard, 2007, 32). Second, you look for that principle or set of principles which, if adopted, would generate as many of these judgements as possible (Pogge, 2007, 163), though you might also consider the relative confidence you have in each judgement, particularly if there is no clear winner from this process (I ignore that issue here). Third, once you have this principle/set-of-principles, you discard whichever judgements are incompatible with it. This means you have 'systematised' your judgements by integrating them into a 'theory' of justice (Pogge, 2007, x, 131, 169). Or, put differently, you have rendered coherent the 'confused sense of justice' that lies beneath the judgements with which you started (Audard, 2007, 32), by aligning both your judgements with each other and the principles to which they (mostly) tend, at which point you have achieved *narrow* reflective equilibrium (Lehning, 2009, 123). Fourth, you examine as many leading moral and political theories as possible, in order to see if any further refinement can be achieved, and by doing so achieve *wide* reflective-equilibrium (Pogge, 2007, 165). Fifth, if everybody in society achieves wide reflective equilibrium on the same set of principles, then we have achieved, not just *general* reflective-equilibrium, given that we agree, but also *full* reflective equilibrium, given that we agree for the right reasons – that is, in virtue of our *wide* reflective equilibriums (Rawls, 2001, 30-31; Lehning, 2009, 124-125; Reidy, 2014, 23).

And then? At that point you are finished, for there is 'no further test available' (Reidy, 2014, 23). Once we go through the first four stages as individuals, and the fifth as a society, there is nowhere else to turn. Or is that misleading? In particular, one might object to this account by questioning the continuity of that fifth stage across Rawls' works as a whole, given his claim in *Theory* that he 'shall not even ask whether the principles that characterise one person's considered judgments are the same as those that characterise another's', and also that 'the views of the reader and the author are the only ones that count' (1971, 50)[7]. To some extent this is settled by the intervening claim that he will simply 'take for granted that these principles are either approximatively the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments divide along a few main lines represented by the family of traditional doctrines that I will discuss' (*ibid*, 50). So, although he does not *ask*, he certainly *assumes*. But there is still a nagging worry. Are *general* and *full* reflective equilibrium really part of his project in *Theory*, or did they only properly emerge later? (And if so, what would that tell us?).

The answer is that that these two ideas – general and full reflective equilibrium - were always there, given the central connection he always drew between justification and agreement. They are there in his claim that it seems '*generally acceptable* that no one should be advantaged […] by natural fortune or social circumstances in the choice of principles [and also] *widely agreed* that it should be impossible to tailor principles to the circumstances of one's case' (Rawls, 1971, 18, emphasis added). They are there in his claim that the 'original position' involves 'conditions that are widely recognised as fitting to impose on the adoption of moral principles' (*ibid*, 584), as well as his admission that it would fail unless 'its conditions are in fact widely recognised, or can become so' (*ibid*, 585). They are there in his early ideal of a 'well-ordered society', given the way in which that ideal connects the goal of justifying principles with the goal of achieving agreement upon them (*ibid*, 453-455), but also his later idea of 'reasonableness', according to which terms of social cooperation can only be fair if they are reasonably acceptable to those to whom they are offered (Rawls, 2005, xlii). And

they are there, most explicitly of all, in his sustained 'remarks on justification' at the close of *Theory*, where he talks explicitly about the dependence of *justification* upon *consensus* on premises, and thus the need for all his key ideas and conditions to be widely shared (Rawls, 1971, 577-587). The talk was always of what 'we' think; of 'our' moral development and psychology; and most obviously of a 'theory' of justice, rather than just a 'distillation of John Rawls' views on justice'. At all times, Rawls wanted agreement on justice amongst at least reasonable people, which is why, although he regretted not distinguishing between different kinds of reflective equilibrium in *Theory* (Rawls, 2001, 31), he was always in pursuit of more than just agreement between himself and a single reader.

But then, if that is right, where do other and sometimes better-known justificatory ideas, such as the 'original position', come in? Here the key is Rawls' claim that the 'original position' is a device of 'representation' (2001, 80-81). It *represents*, by *expressing*, some of the considered judgements of injustice we already have, including our judgement that someone's wealth or power should not affect their choice of principles. As a result, its initial value is simply that it provides an accurate representation of our pre-existing considered judgements (Rawls, 1971, 585), though its ultimate value is a combination of that and our assessment of the principles that result from this situation. So, once again, the basic idea is simple: thoughts are sovereign, in the sense that both principles and the hypothetical procedures by which they are reached are justified (or not) by the extent to which they cohere with the convictions we already have.

Yet surely this justificatory story fundamentally changes between *Theory* and *Political Liberalism*, given the problem of 'stability' Rawls says the latter is intended to address, together with the concepts of 'legitimacy', 'overlapping-consensus', and 'public-reason' he combines to that end (cf. Barry, 1995 with Rorty, 1991)? It does not. Just like the 'original position', such ideas flow naturally from the idea of *full* equilibrium, because with that idea you are already concerned with stability and the shared ideas of our 'public political culture', due to its entailment of *strong* and *widespread* agreement on principles (Pogge, 2007, 165-173). And again: was this concern already there in *Theory*? Yes, because, as explained, Rawls already anchored justification in agreement-on-premises there as a matter of method, whilst simultaneously committing himself to the ideal of a 'well-ordered society' – a society that achieves stable agreement on principles of justice. This crucially requires principles that rest on shared and stable premises, and thus premises that have been chosen from amongst the complete public stock of ideas, given that if our principles rest upon just a few contemplated ideas, they may well change when new ideas come to our attention – hence why *full* reflective equilibrium involves the achievement of *wide* reflective equilibrium by everyone[8]. This is why, as Lehning puts it, notions of public reconciliation, self-clarification, and drawing on a 'shared fund' of ideas are 'recurring elements' right from the beginning (2009, 27-34).

But then what does change? What changes is Rawls' expectation of this process, given that he eventually thought the list of moral and political principles we could agree upon under full equilibrium is shorter than initially imagined, due to differences in both those starting judgements and the wider moral conceptions we already hold. Or, more precisely, he came to believe that liberal conditions, combined with what he calls 'the burdens of judgement', lead to a plurality of 'reasonable comprehensive doctrines' (e.g. Kantianism or utilitarianism), and thus prevent widespread agreement on every aspect of morality, even though we can agree on principles of justice (Rawls, 2005, esp. 54-66). This is why, as Audard puts it, instead of a fundamental 'shift' between *A Theory of Justice* and *Political Liberalism*, he simply becomes

more 'realistic' about the limits of full reflective equilibrium, due to the scale of potential principled agreement (2007, 9, 18, 169-173, 279).

So, justification via reflective equilibrium remains Rawls' core method across all of his key works, as well as the key to every other justificatory device (Freeman, 2007, xiii, 28-29), which is why those works should be understood as a 'coherent whole', right from the beginning (Lehning, 2009, 10), and perhaps even the *very beginning*, according to Laden and Reidy, for whom it is already present in 1946, given the target of an 'explication' of our convictions that would be both stable over time and capable of solving disputes in the real world (2014, 61-62; 2014, 12-13, 27-28). When, therefore, we talk of 'Rawls's method', we should see that it is this continuous idea on which everything else depends.

The empiricism of this method

We turn now to the more general character of justification via reflective equilibrium, and in particular its unappreciated *empiricism*. Consider here (1) the way in which Rawls aligns our 'sense of justice' with our sense of grammar[9] (Rawls, 1971, 47; Audard, 2007, 33; Freeman, 2007, 34; Maffetone, 2010, 145; Pogge, 2007, 164); (2) the way in which he interprets our 'moral capacity' (Rawls, 1971, 46; Freeman, 2007, 37-38), understood as an aspect of our psychology (Rawls, 1999, 289-290; Graham, 2007, 130); (3) the way in which considered judgements function as empirical 'facts' (Rawls, 1971, 51; Freeman, 2007, 31; Laden, 2014, 130; Lehning, 2009, 14; Reidy, 2014, 20) that can be used 'scientifically' (Audard, 2007, 130-131) – and with reference to Goodman's work on induction (Audard, 2007, 33; Maffetone, 2010, 143) – in order to both 'falsify' flawed 'theories' of justice and 'discover' better ones (Audard, 2007, 130-131); and (4) the way those theories both 'explain' the judgements we already have and 'predict' the new ones we ought to adopt (Rawls, 1971, 104, 425; Reidy, 2014, 12-14). What does all this show us? Taken together, it illustrates the fundamentally empirical character of justification via reflective equilibrium, by illustrating the way it turns patterns amongst the thoughts in our heads into principles regarding the way we ought to live together.

More broadly speaking, the method is empiricist (1) in the sense that the justification of principles stems from facts about the world, and in this case the thoughts within our heads (our considered judgements), and (2) in contrast to rationalism, according to which principles are justified as entailments of rationality - hence why Charles Taylor described Rawls' 'aim' as being 'to re-edit something of the Kantian theory, without the metaphysics' (1997, 174). Yet this often goes unnoticed, and for at least two reasons. First, Rawls is not interested in the Platonic *truth* about justice, but simply the principles on which we *already* though only *latently* agree, rightly or wrongly. As Pogge explains, Rawls 'leaves open' the question of correctness, and focuses exclusively on solving 'a practical problem' – that of uniting us on principles of justice by studying our underlying convictions (2007, 163). His 'ambition' is thus not 'to dictate new norms', but simply to 'clarify' our 'common-sense intuitions and beliefs' (Audard, 2007, 14) - or what Laden calls our 'collective sense of right and wrong' (2014, 129). Second, every judgement is *potentially revisable* (Laden, 2014, 130; Maffetone, 2010, 152). This means that any judgement that is incompatible with the principles that cover the maximum possible number of our judgements has to be jettisoned – hence the phrase '*provisional* fixed points' (Rawls, 1971, 20)[10].

Clearly, these two points – the lack of interest in Platonic truths and the revisability of judgements - distinguish Rawls' project from conventional natural science, yet they do so

without leaving empiricism behind. Consider, for example, Rawls' claim that there 'is a definite if limited class of facts against which conjectured principles can be checked, namely, our considered judgements in reflective equilibrium' (Rawls, 1971, 51). Whilst this avoids commitment to ambitions of 'truth', 'proof', or 'correctness', it does commit him to the idea that the only material we have to work with is empirical data taken from the contents of our heads. Yes, he is aspiring to more than just agreement between 'reader' and 'author', and of course he aims for a 'theory' of justice, but it is a theory in the sense of being a distillation and tidying-up of what *we* currently think about justice. It is not a Platonic truth regarding what justice is for all beings in all places and at all times. This is why Lehning is right to talk of 'empirical evidence' being used in the pursuit of 'intersubjective agreement' (2009, 14). Or, put differently, this is why Rawls, together with those who follow him, only tells us what we *should* think about justice on the basis of things we *already* think (Floyd, forthcoming-b; Maffetone, 2010, 147-148). Much against the grain of traditional 'analytic' philosophy, given its perception of a naturalistic 'fallacy', he is going from an introspective 'is' to a political 'ought', or, put differently, from mental facts to political principles.

Yet we need to be careful here, given that saying that Rawls' enterprise is fundamentally empiricist is to say nothing of our *reasons* for adopting that enterprise. In particular we might ask: Is justification via reflective equilibrium driven by simple empirical curiosity or by a deeper and unexamined moral principle? This question matters because it helps us assess the extent of Rawls' empiricism, given that it might transpire that one would *only* ever adopt such an enterprise if driven by a *particular* moral principle. If so, and bearing in mind Cohen's controversial thesis regarding the justificatory relationship between facts and principles (Cohen, 2008, 229-273; for critique, see Jubb, 2009; and Pogge, 2008), we might end up saying, yes, this is an empirical enterprise, but also one that expresses a methodological principle resting on a moral principle. The former, roughly, would be something like 'systematise your judgements into principles and align those principles with those of your fellow citizens'. But what would the latter be?

This point remains under-explored in Rawls scholarship, though there are some interesting – if under-developed – suggestions in these books. Pogge, for example, says the drive towards wide equilibrium might be motivated by the desire to eliminate gaps and contradictions in one's own judgements (2007, 162), whilst the drive towards full equilibrium would need something else, e.g. a desire for principles that can be justified to others (*ibid*, 167; but also Maffetone, 2010, 155). Freeman, alternatively, suggests that the process is driven by Kantian moral autonomy, given that it involves 'reason giving principles to itself out of its own resources' (2007, 38), whilst Audard claims that it expresses both 'autonomy' and 'respect for persons', and notes that the idea of any moral doctrine playing a *foundational* role (as opposed to a *revisable* role via the coherentist workings of the method) would contradict autonomy by imposing values on a person that are not fully their own (2007, 7-8, 17). Laden, finally, claims that Rawls' method involves treating others with 'recognition and respect' (2014, 64), and thus that his way of doing moral and political philosophy is essentially 'moral all the way down' (*ibid*, 65).

One very interesting thing about these interpretations, each of which is plausible, is that in each case morality is not just *independent* of meta-ethics, as Rawls once argued (1999, 286-302), but effectively *trumps* it, given that the way in which we view and come to obtain our moral 'knowledge' must *itself* be compatible with certain values. But again, we need to be careful. Although this is a striking possibility, its significance is somewhat diluted by three further points. First, if these interpretations are treated as accounts of Rawls' motivation in

pursuing wide or full reflective equilibrium, then they are unverifiable, short of some biographical statement from him to that effect. Second, if they are treated as descriptions of an aspect or feature of that theory, then they are mostly compatible with one another - one might simultaneously pursue consistency, respect, and autonomy. Third, if it is possible to pursue wide or full reflective equilibrium on the basis of different principles, then it cannot really be 'driven' by any *one* principle in the relevant sense. It therefore seems dubious to suggest that the method *depends* on any particular moral commitment.

Yet perhaps there is a better interpretation available. Think here of Rawls' claim that 'the correct regulative principle for anything depends on the nature of that thing' (1971, 29). This might suggest that, instead of Rawls' method resting on a prior *moral* principle, it is simply a necessary response to facts about the kind of subject matter a given set of political principles is supposed to 'regulate' (Ripstein, 2010, 684). As a result, although this response can itself be formulated as a principle – 'only pursue principles that are appropriate for the thing they are intended to regulate' – it looks more like a value-free methodological principle than the various moral commitments canvassed above[11]. Consider, for example, how Rawls' principles of justice are intended to regulate a particular kind of society, as defined by things like the 'circumstances of justice' (scarcity and limited altruism), the 'burdens of judgement', the self-identification of individuals in liberal societies as 'free and equal', and the inevitability of a plurality of popular yet reasonable comprehensive doctrines. And consider what Rawls says about the distinctiveness of political philosophy under liberal-democratic conditions (2007, 1-13). With all this in mind, perhaps we could say that, because political philosophy, in the hands of Rawls and his followers, *is* pursued in such a context, as defined by these local values and limitations, it *necessarily* has the character I have described for it under the idea of justification via reflective equilibrium, without having to convey any kind of context-free values or principles?

Again, this is a plausible view, though I think there is still a better way of understanding the morality/principles/values behind the method, bearing in mind, as noted, that it might be possible to 'pursue' it on the basis of *various* principles. Consider the following two scenarios. First, because you desire consistency in your practical-reasoning, or perhaps simply because you are bored, you try to work towards wide reflective equilibrium, and by doing so discover that you are committed to having principles that are justifiable to others, and thus by entailment the idea of full reflective equilibrium, together with the ideas of legitimacy and stability (etc.) it inspires. Second, you already desire to have principles that are justifiable to others, and for that reason work towards wide and then full reflective-equilibrium. What do these two scenarios tell us? They tell us that the motivation for initiating *wide* reflective equilibrium might only be *moral* for *some* of us, even if the later move towards *full* equilibrium does require some such thing. The whole process can be *initially* pursued of a commitment to justice/legitimacy/autonomy/respect/etc. *or* detached academic interest, just as political philosophy in general can begin out of either political conviction *or* philosophical curiosity. All of which amounts to this: justification via reflective equilibrium is an empirical project, the moral credentials of which vary according to the reasons animating whoever pursues it (with such 'reasons' including both the reasons for pursuing wide reflective equilibrium, and the reasons for moving on to full) . And is that a problem for Rawls? Not necessarily, though something else hinted at by these scenarios might be: the possibility that different people, given different starting points, could end up with different equilibriums. It is to three worries related to that possibility that we now turn.

The problems of this method

Consider first of all the worry of the *moral realist*. From their perspective, Rawls' method, given its exclusive focus on pre-existing judgements, seems to rest principles on *opinions* rather than *truths* (as indicated earlier). After all, shouldn't political philosophy tell us what we *should* think about justice, rather than simply re-organising what we *already* think? Rawls' interpreters offer various defences against that charge, including Pogge's claim that moral realism offers no genuine *alternative* to Rawls' approach, given that it cannot help using justification via reflective equilibrium itself in order to discover/justify principles (2007, 176). Similarly, one could argue that Rawls simply *avoids* meta-ethics, as noted, by leaving the question of truth 'open', and instead trying only to solve a 'practical problem' (Rawls, 2005, 110-116; Pogge, 2007, 163, 174-175)[12]. On this view, his work is political, not epistemological (Lehning, 2009, 100); it involves 'public justification', not 'pure demonstration' (Maffetone, 2010, 19). Rather than claiming any special access to Platonic truth, he aims only to solve a practical problem that contemporary citizens *already recognise* on the basis of convictions they *already share* (Maffetone, 2010, 12). This is why, as Reidy puts it, there is no further measure of a set of principles beyond the 'allegiance', post-reflection, of the people for whom they are intended (Reidy, 2014, 19-25). Or, as Pogge puts it, why Rawls' principles are meant to appeal, not just to his philosophical colleagues, but also to his fellow citizens as an 'attractive specification of ideas they already hold' (Pogge, 2007, 196). So, although 'the people' *might* be wrong, there is no way of either *knowing* that, or of *justifying* principles, without *exclusive* reference to the judgements they already hold.

That, however, leaves us with a worry about *relativism*: What if 'the people' disagree with each other? This possibility is acknowledged by Rawls' defenders, who admit that he just has to hope that different individuals, given different starting points, won't end up with different reflective equilibriums (Pogge, 2007, 170; Maffetone, 2010, 156-157). As Lehning explains, 'a minimum [...] (moral) consensus has to be present' (2009, 34, 122), or what Laden calls a 'certain consistency' in the initial 'data' (2014, 130; but also Maffetone, 2010, 11-12, 149, 153-155). And is such consistency available? Do 'the people' agree with each other, at least under wide reflective equilibrium, and at least within individual constitutional democracies, on at least principles of justice, or at least a liberal 'family' of such principles (Rawls, 1999, 579-580)? Clearly, that question cannot be answered here[13]. Instead, it will suffice to note two things: (1) Rawls certainly thought *enough* consistency was available at the start (Reidy, 2014, 20), middle (Rawls, 1999, 306) and end of his career (Rawls, 2005, 14-15), though of course the precise *amount* of postulated consistency is reduced in *Political Liberalism*, as noted above; and (2) he only requires enough consistency across our considered judgements – he does not, for example, worry about meta-ethical convictions, for they are not part of the process.

Even if there is sufficient consistency across our considered judgements, however, that still leaves a third worry – *conservatism*. Consider two possibilities: First, that if we had worked towards full reflective equilibrium in the distant past, we might have affirmed principles that permitted the kinds of discrimination we now think are forms of injustice; second, that by basing principles on our contemporary judgements, we might rule out principles which, in the future, we come to see as progress. Yet this past/future problem reveals a deeper importance to the point that Rawls is pursuing reflective equilibrium in a 'secular, democratic, and scientific age' (Freeman, 2007, x). In short, because such conditions better lend themselves to *considered* judgements and optimally *wide* reflective equilibria, democratic citizens can be treated as 'experts', relative to the subjects of empire or monarchy (Audard, 2007, 7-11). Rawls could then rebut the charge of conservatism by admitting that we should not trust

'data' reached under the wrong conditions, whilst at the same time providing a theory of what the *right* conditions are (Reidy, 2014, 20). And that is important. It means that we can talk, not just of the 'independence' of moral theory, or even of the 'priority' of moral theory over meta-ethics, but also of the 'methodological priority [of] political philosophy over moral philosophy' (Reidy, 2014, 21; but also Freeman, 2007, 284, 310).

This argument, I think, is a powerful one, given that it also helps with both the moral-realist and relativist worries: the former by providing this method with a kind of 'error theory'[14] – an account of why we should grant the judgements of democratic citizens a superior initial credibility – and the latter by, according to Maffetone, providing reason to think that conflicting equilibriums are less likely under the right historical conditions (2010, 145). Yet prioritising political philosophy, in the manner described, still leaves the second half of the conservativism worry standing - the thought that justification via reflective equilibrium, by systematising *today's* judgements, is inimical to *progress*. So what to do? One option, clearly, is to stress the progressive potential of every judgement, as noted, being open to revision (Laden, 2014, 69). Yet that is still *only* potential, for we might just as well end up, as noted, with different equilibriums for different individuals. And that is just the tip of the iceberg when it comes to doubts about our current judgements. Consider the following possibilities: (1) What if no *existing* society is 'just' enough to facilitate the 'right' kind of judgements; and (2) what if we could only trust our theory of what the right kind of society is if that theory rested itself upon judgements reached *within* the right society?

It is with these last points in mind that the final redoubt of Rawls' method emerges. As implied by Pogge's reply to the moral realist, the best thing to say, it seems, when faced with such scepticism, is that there is simply no *alternative* way of doing political philosophy (2007, 176; but also Freeman, 2007, 35, 312). And indeed, perhaps we should not even be that *worried* until we are provided with 'a convincing explanation of why our existing judgements are, *en masse*, unreliable' (Freeman, 2007, 35). The ultimate defence of Rawls' method, therefore, is that unless we can construct an alternative, together with a convincing argument regarding its superiority, we should just 'keep calm and carry on'.

The future of this method

The claim, however, that we should not worry about justification via reflective equilibrium unless our judgements are, *en masse*, unreliable, tells us nothing about what to do if just *some* judgements are problematic. This is why, even if alternatives are unavailable, we might still be interested, in the future, in *amendments* to the details of the method. Consider here, for example, that some philosophers, such as Singer and Parfit, share evolutionary reasons for thinking that at least some of our judgements are unsound, given the way in which they have been 'generated', whilst many Marxists view judgements reached under capitalism as ideological corruptions (Kahane, 2011). As a result, we might want to *eliminate* some of the thoughts we currently enter into this process, or even most of those thoughts, in which case we would be using a minority of our judgements to outweigh the majority, rather than *vice versa*.

Similarly, we might want to change the way we *weigh* our judgements. Bearing in mind the shared interest of the recent ideal/nonideal-theory and moralism/realism debates in turning principles into practical guidance (Galston, 2010, 392-394; Wiens, 2012), perhaps we should give greater weight to judgements about specifically *political* practices (e.g. 'racial disenfranchisement is wrong') relative to abstract intuitions about things that are only

relevant via loose analogy (e.g. runaway 'trolleys')? Or we could do the complete opposite: Bearing in mind the anti-ideological worries of Marxists and realists, the evolutionary worries of Parfit and Singer, and the conservatism worry detailed above, perhaps we should give greater weight to 'purer' abstract intuitions over 'corrupted' political judgements? Admittedly, even those thoughts might be too ideological for some Marxists, yet even so, it might still be better to work with our intuitions in response to, say, hypothetical camping trips (Cohen, 2009), if our *only* choice is between these and our 'corrupted' judgements about real-world cases of justice/injustice.

Clearly, the only solid upshot of these possibilities is that we need to think more in the future about both the different types of thought we enter into this method and their relative merits. Though note the mildness of that claim. We are only contemplating *revisions* to justification via reflective equilibrium in the sense of changing its inputs (so one might even say that this would not change it all, but would only change our theory of 'considered judgements'). We are not, for example, contemplating the 'continental' view, according to which every judgement is merely a construction open to re/de-construction (Floyd, forthcoming-a), or the unlikely idea of deriving political principles, not from patterns in the way that we *think*, but from patterns in the way that people *behave* in response to different political conditions (Floyd, 2011). And that, I think, is unsurprising. Given how much Rawls has standardised this method, as these books attest, it is hard to imagine jettisoning it in the way that might be required by either of those approaches.

Yet note, finally, that *standardising* this method does not mean inventing it *ex nihilo*, which is why a further issue for its future concerns the historical question of its *uniqueness*. How much, for example, does justification via reflective equilibrium differ from Plato's dialogues, given the way they obliterate our initial judgements of justice/injustice? How much does it borrow from Aristotle or, more pertinently, Sidgwick, given that Rawls explicitly identifies his method with *both* of their approaches? (1971, 51) Whatever the answers, given the need to contemplate both amendments and alternatives to this method, it cannot be long before a concern with methodology in political philosophy leads to a greater concern with the history of our methods. But that is for another day. For now, Rawls' method remains standard, which is why the principal virtue of these books remains the light they shed on its nature. Or, more generally, the light they shed on *his* legacy and *our* blueprint.

References

Barry B (1990) *Political Argument*. Berkeley, California:University of California Press.

Barry B (1995) John Rawls and the search for stability. *Ethics* 105(4):874-915.

Berlin I (1962) Does Political Theory Still Exist? In: Laslett P & Runciman WG (eds) *Philosophy Politics and Society*. Oxford:Blackwell, pp.1-33.

Cohen GA (2008) *Rescuing Justice and Equality*. Cambridge, MA:Harvard University Press.

Cohen GA (2009) Why not socialism? Princeton, NJ:Princeton University Press.

Floyd J (2011) From historical contextualism, to mentalism, to behaviourism. In: Floyd J & Stears M (eds) *Political Philosophy versus History? Contextualism and Real Politics in Contemporary Political Thought*. Cambridge:CUP, pp.38-64.

Floyd J (forthcoming-a) Analytics and Continentals: divided by nature but united by praxis? *European Journal of Political Theory*.

Floyd J (forthcoming-b) *Is political philosophy impossible?* Cambridge:CUP.

Galston W (2010) Realism in Political Theory. *European Journal of Political Theory* 9(4), 385-411.

James A (2012) *Fairness in Practice*. Oxford:OUP.

Jubb R. (2009) Logical and epistemic foundationalism about grounding: The triviality of facts and principles. *Res Publica* 15(4):337-353.

Kahane G. (2011) Evolutionary debunking arguments. *Noûs 45*(1):103-125.

Laden AS (2014) Constructivism as Rhetoric. In: Mandle J & Reidy DA (eds) *A Companion to Rawls*. Chichester:WileyBlackwell, 60.

Laslett P (1956) *Philosophy, Politics and Society*. Oxford:Blackwell.

Pogge T (2008) Cohen to the Rescue! *Ratio* 21(4), 454-475.

Rawls J (1971) *A Theory of Justice*. Cambridge, MA:Harvard University Press.

Rawls J (1999) *Collected Papers*. Cambridge, MA:Harvard University Press.

Rawls J (2005) *Political Liberalism*. New York, NY:Columbia University Press.

Rawls J (2007) *Lectures on the History of Political Philosophy*. Cambridge, MA:Harvard University Press.

Reidy DA (2014) From Philosophical Theology to Democratic Theory. In: Mandle J & Reidy DA (eds) *A Companion to Rawls*. Chichester:WileyBlackwell, pp.9-30.

Ripstein A (2010) Critical Notice. *Canadian Journal of Philosophy*, 40(4):669-699.

Rorty R (1991) *Objectivity, Relativism, and Truth*. Cambridge:CUP.

Sayre-McCord G (1996) Coherentist Epistemology and Moral Theory. In Sinnott-Armstrong W & Timmons M (eds) Moral Knowledge? Oxford:OUP, pp.137-189.

Taylor C (1997) Leading a Life. In: Chang R (ed) *Incommensurability, Incomparability and Practical Reason*. Cambridge, MA:Harvard University Press, 170-183.

Wiens D (2012) Prescribing Institutions without Ideal Theory. *Journal of Political Philosophy*, 20(1):45-70.

Williams B (2005) *In the Beginning was the Deed*. Princeton, NJ:Princeton University Press.

Wolff J (2013) Analytic Political Philosophy. In: Beaney M (ed) *The Oxford Handbook of the History of Analytic Political Philosophy*. Oxford:OUP, pp.795-824.

[1] For alternative accounts, see (Barry, 1990) and (Wolff, 2011).
[2] The key here is 'principled' recommendations. Although Berlin highlights *conflicts* between values, and occasionally hints at how they should be *weighted*, he doesn't propose principles and priority rules for such trade-offs in the manner that Rawls and his followers do. I thank Rob Jubb for pressing me on this point.
[3] For an alternative case that constructivism is subordinate to reflective-equilibrium, see (Laden, 2014, 62-65).
[4] Note though that reflective equilibrium and coherentism are often treated as synonymous. See Sayre-McCord.
[5] As to whether this 'new' era was the 'first' such era – I comment on this at the end of the article. For further discussion of 'thoughts-to-oughts', see (Floyd, forthcoming-b).
[6] Note that these stages can be run in a slightly different order, provided that all judgements/principles/traditions are included (Pogge, 2007, 28, 162-170).
[7] I am grateful to an anonymous reviewer for this journal, and to Rob Jubb, for pressing me on this point.
[8] Note that this is without even considering the stability requirement within the original position, as described in *Theory*.
[9] One might object here by noting, as Rawls does, that theories of grammar can outrun and overrule the 'data' from which they start. This is true, but only to a small extent. Just as we would not accept a theory of grammar that radically revises conventional language usage, so Rawls does not want (or consider justifiable) a theory of justice that rejects most of our current thinking on the subject.
[10] Though note that some points seem anything but provisional, e.g. Rawls' faith in the wrongness of slavery.
[11] This issue is central to recent discussions of 'practice-dependence', e.g. (James, 2012)
[12] This becomes ironic if reflective equilibrium is (a) re-labelled as constructivism and (b) treated as itself a meta-ethical theory, though it need not be so treated. See (Laden, 2014, 59).
[13] Though see (Floyd, forthcoming-b)
[14] Such an 'error theory' is precisely what Williams thought liberalism lacked, though I cannot explore its plausibility here (Williams, 2005, 66—67).