

Sidgwick, Reflective Equilibrium and the Triviality Charge

Michael W. Schmidt, Karlsruhe Institute of Technology, Germany

Abstract

I argue against the claim that it is trivial to state that Sidgwick used the method of wide reflective equilibrium. This claim is based on what could be called the Triviality Charge, which is pressed against the method of wide reflective equilibrium by Peter Singer. According to this charge, there is no alternative to using the method if it is interpreted as involving all relevant philosophical background arguments. The main argument against the Triviality Charge is that although the method of wide reflective equilibrium is compatible with coherentism (understood as a form of weak foundationalism) as well as moderate foundationalism, it is not compatible with strong foundationalism. Hence, the claim that a philosopher uses the method of wide reflective equilibrium is informative. In particular, this is true with regard to Sidgwick.

Introduction

This paper contributes to the debate whether Sidgwick used the method of reflective equilibrium. “Reflective equilibrium” is the name of the method of justification which John Rawls suggests in his *A Theory of Justice* (1971). He claims that many other philosophers have used the method even before: Most prominently Nelson Goodman in *Fact, Fiction, and Forecast* (1955) and Henry Sidgwick in *The Methods of Ethics* (1907). Whereas Catherine Elgin – a scholar sympathetic to Goodman – endorsed Rawls’s suggestion and even worked out a better understanding of the methodology, from the very beginning there was a dispute over the claim that Sidgwick employed such a method. I argue that Rawls’s claim that Sidgwick used the method can be – in some sense – defended against certain strong criticisms.

The following is divided into three parts. In the first part of this paper, I will provide some important background information: I will do that by referring to an article by Peter Singer published 1974 entitled “Sidgwick and Reflective Equilibrium” in which he argued that reflective equilibrium is neither an adequate method of justification nor the method Sidgwick did employ. I will reconstruct Singer’s interpretation of Sidgwick and his argumentation against Rawls’s claim.

This sets the ground for the second part. I will begin by sketching the reasons why Singer slightly revised his earlier criticism of the method of reflective equilibrium in his latest works. After considering the now predominant wide interpretation of the method Singer

now holds that it is indeed possible to claim that Sidgwick used it – but he still would refrain from saying so, because of one more charge he presses against the method and the claim: the charge that if it is understood in the wide sense, it is simply trivial to state that a philosopher makes use of it. According to the charge there is no alternative to the method of reflective equilibrium, if it is interpreted in such a wide way that its use involves all relevant philosophical background arguments, because then it includes all other rival methods. So, to state that someone uses the method settles nothing and is pointless. This is what I dubbed the *Triviality Charge*.¹

In part three I will assume that the wide interpretation of the method of reflective equilibrium, which Singer considers compatible with the method of Sidgwick, is the only plausible interpretation. I will argue that it is – even in the wide interpretation – not trivial to state that some philosopher and especially that Sidgwick did use it. Hence, by refuting the claim of the *Triviality Charge*, I will argue that it is informative and justified to state that Sidgwick used the method of reflective equilibrium, even if one takes granted that Singers interpretation of Sidgwick's *Methods of Ethics* is the correct one or grasps the important methodological points adequately.

I Different Interpretations: Sidgwick and Reflective Equilibrium

According to Singer's interpretation in "Sidgwick and Reflective Equilibrium", Sidgwick proposes a top-down approach in the realm of normative ethics to justify moral propositions: one has to start with self-evident axioms (in the form of universal principles) and to see what follows from these. If our everyday moral judgments cohere with the ethical theories derived from the axioms, this can be used as an argument to convince common people to adopt the justified ethical theories – but this doesn't show the justification of these theories, since they themselves are only justified by their status as being inferentially connected to the self-evident axioms, which have a privileged epistemic status. Yet, still one can be mistaken in holding an apparent axiom to be a real axiom. In terms of Laurence Bonjour widely used in epistemology one can call this a moderate foundationalism, where some basic beliefs – here the self-evident axioms – are themselves justified without being inferred

¹ I am borrowing the name "Triviality Charge" from Julia Langkau. Langkau argues contra the charge against reflective equilibrium in a different (non-moral but epistemic) context and takes a different line of argument, though one could say, we share the same strategy. Cf. Langkau 2013.

from other beliefs. They can pass on by inference relation the justification to other beliefs and this suffices, given that we have true beliefs, that we also have knowledge – in our case moral knowledge (cf. Singer 1974, 498-501, 503-5, 507-8; cf. BonJour 1985, 26-30).

So, according to Singer, the basic beliefs in Sidgwick's moderate foundationalism are on the most abstract level of moral entities: Sidgwick's basic beliefs concern ultimate ethical principles – namely the *principle of justice*, the *principle of prudence* and the *principle of benevolence* – from which the morally right theories are deducible. So, in order to choose the right moral theory, it is essential to have an intuitive insight in the self-evidence of the axioms and to check if the self-evidence was merely apparent self-evidence by reflecting if there are any other self-evident axioms that conflict with the one under investigation, if there is an consensus on the axioms and if the principles corresponding to the axioms are ambivalent or precise. As I see it, this interpretation of Sidgwick – which I tried to reconstruct here in a condensed form – remains Singer's interpretation and hasn't changed substantially in the other works I will refer to (cf. Singer 1974, 503, 507-8; cf. Sidgwick 1907, Book III, esp. Chapter XI, 2, and Chapter XIII).²

He contrasts this moderate foundationalism in the realm of normative ethics with an interpretation of the method of reflective equilibrium. The basic idea of reflective equilibrium is that a theory and our common sense considered judgments should be brought into agreement. And if they both support each other in the best available way the judgments as well as the corresponding theory are justified. Both are also open to revision in the process of adjustment.

Although sometimes the method of reflective equilibrium is accused of being a form of disguised common sense-intuitionism – which means that it is a bottom-up moderate foundationalist approach that presupposes that one can do ethics analogue to (some common interpretations of) empirical inquiry or science – mostly it is recognized as a form of coherentism, as it is by Singer. Importantly Singer seems to imply that the use of the method of reflective equilibrium would result not only in a coherence account of justification but also a coherence account of validity or truth (cf. Singer 1974, 492-5, esp. 493-4).

Rawls, according to Singer, thus misinterprets Sidgwick when he suggests that they share the same method of justification. He thinks that this misinterpretation rests on the passages where Sidgwick tries to show that the utilitarian theory, which can be derived from the

² There are, of course, other interpretations: Rawls himself refers to Schneewind 1963. Skelton 2010 backs some of the points that lead to a rejection of the claim that Sidgwick used the method of reflective equilibrium, Crisp 2002 is as well critical on the suggestion that he used this method, but on different grounds, Sverdlik 1985 and Brink 1994 have interpretations that would in contrary back the claim of the direct use of the method.

principle of benevolence together with *the principle of justice*, is fitting best to the judgments of our common-sense morality. But this *ad hominem* argument – as Singer calls it – (respectively the coherence with our common sense judgments) is not what justifies this axiom (or any other). What justifies all possible axioms is, that they are self-evident and remain self-evident after due reflection.

Now if the method of reflective equilibrium were to be understood in the way Singer suggested, and one accepts Singer's interpretation of the *Methods of Ethics* it would clearly be inadequate to state that Sidgwick used the method of reflective equilibrium. Sidgwick, according to Singer, would have used a top-down moderate foundationalism and believed in objective moral truths, whereas Rawls would have used a coherentist approach, that includes not only a coherence account for justification but also for truth and thus he would be a subjectivist or cultural relativist concerning moral truth, so that their methodology is not consistent at all.

II Singer's Revised Position

This brings us to part two of this paper. We can begin by asking the question: Was Singer right with this interpretation of reflective equilibrium? Most often reflective equilibrium is – as Singer rightly suggested – indeed considered a coherentist method of justification. But typically, this involves a form of weak foundationalism, which means that while there are no beliefs which themselves are already justified without inferential backing, there are some which have an initial credibility, because they are what we in fact believe before we start to scrutinize and criticize our system of beliefs. This initial credibility is – according to the weak foundationalist interpretation of coherentism (which I want to presuppose henceforth) – not enough to grant an inquirer knowledge. Justification thus arises only if beliefs – initially held or not – can be incorporated in a system of held beliefs in the most coherent way such that they are mutually backed by inferential relations better than in any alternative system of beliefs that one could accept (cf. Rawls 1974, 8).

But although reflective equilibrium is widely understood as a coherence method of justification, that doesn't mean that a proponent of it must embrace a coherence account of truth: There can be objective moral truth that is not created nor secured by a coherent

system of moral beliefs even if one thinks that this is what justifies these beliefs.³ Most proponents of the method of reflective equilibrium take coherence (combined with initial credibility of the beliefs that we actually accept) as a criterion of justification but not the criterion of truth and thus admit, contrary to Singers earlier assumption, that there is (moral or non-moral) truth or objectivity independent of its subjective justification.

One of the most detailed accounts of reflective equilibrium originates from Norman Daniels (1996). He builds his account on the distinction between a narrow reflective equilibrium and a wide reflective equilibrium.⁴ If one is trying only to achieve a narrow reflective equilibrium one just tries to achieve coherence only between our considered judgments and theories. But according to the method of reflective equilibrium in the wide interpretation – which we should use in philosophy – one has to incorporate in the weighing process all relevant background theories and arguments. These background theories are scientific or philosophic theories or arguments that would have an impact on the narrow reflective equilibrium, were they to be considered.

Thus, the method of reflective equilibrium in the wide sense is a method that enables us to be critical of our judgments and scrutinize our biased system of belief. It is also wide enough for background theories, that mandate for special areas of investigations special sub-methods. If, for example, a plausible theory casts doubt on our common sense moral judgments, it could be possible to discredit these judgments systematically in moral inquiry – just as Singer himself holds – with the possible result (if it also can be argued for the remaining beliefs to be basic) that one establishes for the realm of normative ethics a moderate foundationalism. In this way, wide reflective equilibrium is indeed compatible with the method of Sidgwick the way Singer interpreted it.

Weak foundationalism – according to this interpretation of reflective equilibrium – remains the “default setting” for inquiry unless a different sub-method is vindicated for certain areas of investigation.

Also, Singer himself, who, of course, did follow the debate on reflective equilibrium, now explicitly accepts in his 2005 article “Ethics and Intuitionism” and his book *The Point of View*

³ Indeed, Rawls himself leaves room for the idea that if we use the method of reflective equilibrium it might result in a convergence of our ethical belief systems what could indicate that we are getting closer to moral truths. That implies on the other hand, that we still can go wrong, even if we have reached a reflective equilibrium. Cf. Rawls 1974, 9-10, 21. Cf. Daniels 1996, 33-40.

⁴ According to Rawls' terminology in his article “The Independence of Moral Philosophy” (1974) – Rawls suggested the use of wide reflective equilibrium already in *A Theory of Justice*. Cf. Rawls 1971, 49; 1993, 8-9.

of the Universe – jointly written with Katarzyna de Lazari-Radek (2014) – that the method of wide reflective equilibrium is indeed compatible with Sidgwick’s method.⁵

Admittedly, it is possible to interpret the model of reflective equilibrium so that it takes into account any grounds for objecting to our intuitions, including those that I have put forward. Norman Daniels has argued persuasively for this ‘wide’ interpretation of reflective equilibrium. If the interpretation is truly wide enough to countenance the rejection of all our ordinary moral beliefs, then I have no objection to it. (Singer 2005, 347)⁶

II.a The Triviality Charge

So, what is stopping us – granted that we do agree with Singer’s Sidgwick interpretation – from simply stating that Sidgwick indeed used something like the method of wide reflective equilibrium? According to Singer there is a price to be paid, if the method of reflective equilibrium is understood as wide, as Daniels suggests:

The price for avoiding the inbuilt conservatism of the narrow interpretation, however, is that reflective equilibrium ceases to be a distinctive method of doing normative ethics. Where previously there was a contrast between the method of reflective equilibrium and “foundationalist” attempts to build an ethical system outward from some indubitable starting point, now foundationalism simply becomes the limiting case of a wide reflective equilibrium. (Singer 2005, 347)

Singer claims that if the use of the method of reflective equilibrium is no means anymore to distinguish a priori, that some moral philosopher proposes rather a coherentism than a moderate foundationalism for ethical inquiries, then stating that the philosopher used the method of reflective equilibrium becomes meaningless. It’s just trivial to state, that a philosopher used the method of reflective equilibrium understood in this wide sense, that includes the possibility of an ethical moderate foundationalism – so to say that Sidgwick used the method of reflective equilibrium is pointless and we shouldn’t state pointless utterances (at least, this seems to be implied). This is the core of the triviality charge:

[...] whether wide reflective equilibrium and foundationalism can be distinguished depends on the substance of ‘the acceptable moral theory’ and on what the philosophical arguments allow us to conclude. Without knowing which moral theory is acceptable and whether there are

⁵ He accredits the notion of wide reflective equilibrium to Daniels and not to Rawls – and I think he is mistaken here, Cf. Singer and de Lazari-Radek 2014, 11-114; cf. Singer 2005, 347.

⁶ One could interpret the cited statement in a way that Singer now accepts a weak foundationalism in which certain beliefs are epistemically devalued, but I suggest that Singer’s position is still a moderate foundationalism.

philosophical arguments that reveal which moral judgments are objectively true, we cannot exclude the possibility that, once we have found the soundest moral theory and the best philosophical arguments, we will be able to demonstrate that none, or virtually none, of our existing moral judgments are credible; such that we can confidently reject all, or virtually all, of our current moral judgments, and replace them with the judgments that follow from the moral theory. [...] In that case, the distinction between wide reflective equilibrium and foundationalism has narrowed to a vanishing point. It would then be true, but trivial, that when we do normative ethics, there is no alternative to the method of reflective equilibrium. There would be no alternative because wide reflective equilibrium is so wide that it includes all possible methods, including foundationalism. (Singer and de Lazari-Radek 2014, 112-3)⁷

III Arguments against the Triviality Charge

I will argue that it is not trivial to state that Sidgwick used the method of reflective equilibrium (understood as the attempt to achieve a wide reflective equilibrium).

To support this thesis, I will advance one main argument and two further arguments.

1. The main argument takes that *fallibilism* is a necessary condition for being able to support the method of reflective equilibrium. Fallibilism is roughly defined here as the presupposition that all beliefs without exception which we could use in our philosophical argumentation are in principle questionable and open to scrutiny since the epistemic agent cannot be certain of their truth. Why is fallibilism a necessary condition for the method of reflective equilibrium? If there were some beliefs which were not open to scrutiny, they would be totally fixed points in the process of adjustment when conflicts occur in our system of belief. And everything that could be brought in deductive inferential relation with them were as well as fixed (given that we accept some set of rules of logic as one of these fixed points). But fixed points are not open to weighing considerations if they conflict with other beliefs in a system that is to be brought in a reflective equilibrium. All the balancing process that lies at the core of the process of achieving reflective equilibrium is only possible if we accept fallibilism. Yet, fallibilism isn't the only condition for the method of reflective

⁷ A similar position is held by Sem de Maagt: "One problem is that by including just about any possible disagreement related to the justification of our moral beliefs into its methodology, reflective equilibrium runs the risk of becoming vacuous as a method of moral justification, because ultimately reflective equilibrium will simply be reduced to reasoning about ethics in general. That is, if any kind of disagreement is included in the search for reflective equilibrium it is not evident that it still can function as a method of moral justification." (2017, 458).

equilibrium: A second further necessary condition is a (weak) holistic perspective of justification which includes inferential interdependence. A third condition consist in criteria of rational belief revision in case of inconsistent beliefs in a holistic system of beliefs – namely that a belief succeeds over another conflicting one in the case it has (in light of all supportive evidence) a higher degree of credence than the other belief.⁸ A fourth condition would be the weak foundationalist assumption that all the beliefs we think to be true – the “considered judgments” or “commitments” – have an initial weak credibility that suffices to distinguish them from merely possible beliefs but that is not strong enough for granting us knowledge without further inferential backing.

But to challenge the triviality charge, I will only rely on fallibilism.

Now there are, at least, two reasons why it is informative – rather than trivial – to state that Sidgwick was a fallibilist:

- i. The first reason is obvious: There are philosophers who are infallibilists and think that they have reached some unquestionable truths. Singer himself points to Descartes to illustrate this position from which he wants to distinguish Sidgwick’s position. This means obviously that Singer also thinks that pointing to Sidgwick’s fallibilism is informative. In fact, if one wanted to classify that kind of infallibilism that Descartes seems to present, one would call it – according to BonJour – strong foundationalism.⁹ So it is literally false to simply state that foundationalism might be included in the method of reflective equilibrium: Some form of foundationalism might not be included, which shows us again, that to assign the method of reflective equilibrium to some someone is informative. Granted – nowadays most philosophers seem committed to some sort of fallibilism but there still might be some who argue for infallible truths – in ethics as well as in other areas of philosophy. So even with the majority of contemporary philosophers being fallibilists it is not completely trivial to state that some philosopher is a fallibilist. This already counts for contemporary philosophers, but Sidgwick is a historic philosopher, so the claim that he used the

⁸ This should normally lead to a maximal coherent set of beliefs, what could in itself be counted as another condition for reflective equilibrium.

⁹ One could argue that Descartes only seems to present a strong foundationalism and if we were to reconstruct his position with the principle of charity in mind it would turn out that he too was a fallibilist. This might hold for other philosophers as well. If so, the claim that the method of reflective equilibrium is the method of philosophy could be true (generally speaking) – I leave this open to further investigation. But even if this were correct, it still would be informative to state that these philosophers were fallibilists because one had to debunk their apparent infallibilism and theoretically infallibilism would still be an option.

method of reflective equilibrium is even more informative and interesting, if we rightly can assume that there have been more infallibilists in the past (Crisp 2002, 60-63).

- ii. The second reason is an exegetic one: In Sidgwick's *Methods* the notion of self-evident axioms plays – as it is well known – a fundamental role. But how is the notion of self-evidence understood in Sidgwick's *Methods of Ethics*? To state that the axioms he proposes are based on a self-evident intuition (plus reflection) *but* still are fallible is an informative interpretation.¹⁰

This is my main argument and I would like to turn to two additional arguments against the Triviality Charge

- 2. The second argument concerns the meta-level on which the methodological design of a subdomain, i.e. certain areas of investigation with specific features, is justified. On this meta-level we are operating on coherentist or weak foundationalist standards even if there is a moderate foundationalist standard (or any other fallible standard) justified in the sub-domain. A sub-domain could be for example the area of normative ethics. If one does assume that Sidgwick proposed a moderate foundational method for normative ethics but was using a reflective equilibrium to justify this method on the meta-level of – let's say – metaethics, then on this meta-level he was arguing in a coherentist or weak foundationalist way. I hold that the same pattern is true for externalist epistemologists. Since these claims could be more controversial than the precedent, I would like to build upon it only an additional argument against the *Triviality Charge*. Yet I think, it might be the philosophically more interesting claim. There seem to be – at least – two non-trivial statements connected with this claim:
 - i. The arguments by which a moderate foundationalism in a subdomain like normative ethics is justified are establishing the methodological design of the subdomain in the first place and are thus superordinate. In other words, the “default position” of inquiry is coherentism or weak foundationalism and a change from the default position must be justified and held justified over time in coherentist or weak foundationalist terms.

¹⁰ Singer himself emphasizes that Sidgwick is a fallibilist: Cf. 1974, 508.

- ii. It follows then that the methodological design of a subdomain that changed from the “default position” is open to ongoing scrutiny in the always dynamic remaining process of the method of reflective equilibrium. To change the method in a subdomain is always provisional, as every justification in reflective equilibrium is provisional and open to further scrutiny. Trying to achieve a reflective equilibrium remains always an ongoing task because there always can be new beliefs – for example through new experiences or evidence – that would have to be incorporated in the holistic system of beliefs. And there is always the possibility that one overlooked relevant background theories or arguments. There could be new inconsistencies at any time which we did or could not anticipate, such that trying to achieve a reflective equilibrium is an ongoing dynamic process. It seems that a perfect reflective equilibrium is a philosophical ideal, but even this ideal state – at a certain time – would be provisional.

If this is all true, then it is clearly non-trivial.

3. The third argument points to the fact that fallibilism is an ideal as well as an important attitude in important domains of civil society: for example, in science or in the political sphere of liberal democracies. This renders information about the use of the method of reflective equilibrium informative and valuable in a social sense – it is useful for us as citizens and epistemic agents if it is stressed that an important theory is fallibilistic. As this is a claim that must be explained and argued for in detail (which I cannot do extensively here), I consider it only an additional possible argument, which I want to point to at the end of my argumentation.

IV Conclusion

To conclude: Is it misleading or inadequate to state that Sidgwick used the method of reflective equilibrium as his method of justification? Since the method is capable of justifying in the realm of ethics coherentism as well as moderate foundationalism, it is at least important to qualify how Sidgwick used the method exactly: Did he use it to establish a moderate foundationalism for the area of normative ethics with his abstract universal axioms as fallible basic beliefs? Or did he treat his axioms merely as provisionally fixed points (like Rawls treats the judgment about the injustice of slavery as a provisionally fixed point) but not as basic beliefs in the sense that is needed for a moderate foundationalism? To decide the correct answer is a goal for experts on Sidgwick’s philosophy. So far Singer’s position might be vindicated.

But the claim that he used the method of reflective equilibrium in general as his method of justification gives no preliminary decision to this question and is nevertheless quite informative.¹¹

And as Sidgwick wasn't merely a fallibilist, but also tried to argue in a coherent way for his position with respect to personally held judgments, I would say it is also quite safe to suggest that he made use of the method of reflective equilibrium.

References

- [1] Bonjour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, MA/London: Harvard University Press.
- [2] Brink, David O. 1994. "Common Sense and First Principles in Sidgwick's Methods." *Social Philosophy and Policy* 11 (1): 179-201.
- [3] Crisp, Roger. 2002. "Sidgwick and the Boundaries of Intuitionism." In *Ethical Intuitionism: Re-evaluations*, edited by Philip Stratton-Lake. Oxford: Oxford University Press, 56-75.
- [4] de Maagt, Sem. 2017. "Reflective equilibrium and moral objectivity." *Inquiry: An Interdisciplinary Journal of Philosophy* 60 (5): 443-65.
- [5] Daniels, Norman. 1996. *Justice and Justification*, Cambridge: Cambridge University Press.
- [6] Elgin, Catherine. 1996. *Considered Judgment*, Princeton: Princeton University Press.
- [7] Goodman, Nelson. 1955. *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.
- [8] Langkau, Julia. 2013. "The Method of Reflective Equilibrium and Intuitions" In *Was dürfen wir glauben? Was sollen wir tun? Sektionsbeiträge des achten internationalen Kongresses der Gesellschaft für Analytische Philosophie*, edited by Miguel Hoeltje, et al. University of Duisburg/Essen, 352-64. urn:nbn:de:hbz:464-20130612-081113-3
- [9] Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

¹¹ It would be an interpretative inquiry of its own to determine if this general claim was Rawls's claim or if he indeed interpreted Sidgwick's epistemology for normative ethics as a form of coherentism and pointed therefore to his use of the method of reflective equilibrium.

- [10] _____. 1974. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 48: 5-22.
- [11] _____. 1993. *Political Liberalism*. New York: Columbia University Press.
- [12] Sidgwick, Henry. 1907. *The Methods of Ethics*, 7th edition. London: Macmillan and Co.
- [13] Singer, Peter. 1974. "Sidgwick and Reflective Equilibrium" *The Monist* 58 (3): 490-517.
- [14] _____. 2005. "Ethics and Intuitionism" *The Journal of Ethics* 9 (3-4): 331-52.
- [15] _____, and Katarzyna de Lazari-Radek. 2014. *The Point of View of the Universe. Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.
- [16] Skelton, Anthony. 2010. "Henry Sidgwick's Moral Epistemology" *Journal of the History of Philosophy* 48 (4): 491-519.
- [17] Schneewind, J. B. 1963. "First Principles and Common Sense Morality in Sidgwick's Ethics" *Archiv für Geschichte der Philosophie* 45 (2): 137-56.
- [18] Sverdlik, Steven. 1985. "Sidgwick's Methodology" *Journal of the History of Philosophy* 23 (4): 537-53.