T.R. Edward
10th July 2023

Reflective equilibrium in political philosophy and its critics

What is reflective equilibrium? In the reflective equilibrium method that John Rawls recommends, we evaluate a set of principles regarding social justice (a theory) by comparing it with our considered moral judgments (the data). The theory will usually consist of a few very general principles, such as one or two or three. If there is a conflict with most of our considered moral judgments, then we must change the theory. But if the theory entails most of our considered moral judgments, while conflicting with a few, then there is the option of abandoning these few instead. If at some stage in the process of comparison and revision, we have entailment and no conflict, then we have reached reflective equilibrium. (Note: what has been described is narrow reflective equilibrium. Also: I shall use "fit" to refer to entailment and "lack of fit" to refer to inconsistency; there is something in-between which, for convenience, I shall pretend is not there.)

What is a considered moral judgment? A considered moral judgment, says Rawls, is a judgment made in conditions where one's moral capacities are most likely to be displayed without distortion (1999: 42). He sets aside those made when hesitant, upset, and frightened. Considered moral judgments can include judgments about specific cases (e.g. it was wrong to set fire to his house) and more general judgments (e.g. it is wrong to set fire to a person's house for no reason), but the method is easier to grasp by limiting the focus to the specific. Below I shall sometimes use the term "intuitions" for our considered moral judgments. Note: Georg Brun denies that the use of the term "intuition" here corresponds with standard philosophical usage (2014).

Grouping objections. The objections below are presented in chronological order, but some of the objections are very similar and it is natural to group them together. I generally omit defences.

1. The variation in intuitions objection. There is no common set of intuitions that all will bring to the process. One person will compare their theories with one set of intuitions and another person with another set. The method therefore seems unsuited to achieving Rawls's aspiration of consensus on one theory of justice, even among "reasonable citizens."

2. The under-determination objection. Assume, for the sake of argument, that we do all bring the same set of intuitions to the process. Still, reflective equilibrium allows for multiple theories that are equally justified, e.g. if theory 1 fits with all but one intuition and theory 2 fits with all but one intuition, a different one, and no other desirable quality decides between them (e.g. simplicity). Neither theory is better for achieving equilibrium. But, as mentioned before, Rawls aspires to consensus.

3. A descriptive method. Reflective equilibrium seems a method more suitable for describing a moral sensibility, as an anthropologist might, rather than working out what is true – what is the morally right thing to do – because the principles merely provide a systematic model for deducing a person's intuitions. (This objection does not fit well with the option of abandoning intuitions.)

4, Overly conservative. Reflective equilibrium is overly conservative because it accords such a large role to the intuitions an individual has to begin with, testing proposed principles by how well they fit with these, and making the conditions on when one can abandon an intuition which conflicts with principles demanding.

5. Objection from Peter Singer. If it can be shown that a certain moral theory follows from a self-evident proposition or set of self-evident propositions, then we should regard it as a true theory (1974: 516). It should not be compared with our moral intuitions in order to potentially lead to revising the theory, because the starting point for deducing the theory was a self-evident proposition or set of them, so the theory must be true if one has reasoned correctly. The number of intuitions it conflicts with does not matter. Thus if we proceed in this way, we should reject reflective equilibrium as a method. The utilitarian Henry Sidgwick tries to proceed in this way, Singer says and seems to endorse this approach himself.

6. Objection from Richard B. Brandt. The reflective equilibrium procedure lacks justification. If we assume that principles have implications for action, then that brings a requirement of consistency, because it would be self-defeating to espouse inconsistent principles, says Brandt, but he tells us that this does not lead to a more specific procedure which involves instructions regarding how to deal with conflicts between principles and intuitions (1979: 20).

7. Another objection from Richard B. Brandt. Reflective equilibrium is related to coherentist rather than foundationalist theories of knowledge – theories which specify a certain kind of belief or other entity as the foundation for knowledge, e.g. self-evident beliefs or sensations. For Brandt, it suffers from an analogous worry to earlier coherentist theories, which say that coherent belief systems are justified – the system could be coherent yet a fiction. There needs to be good reason to regard our moral intuitions, which the procedure directs us to largely achieve coherence with, as true and Brandt thinks there is no good reason. This objection seems more discussed (see Scanlon 2003: 146; Tersman 2018).

8. Objection from Joseph Raz. In reflective equilibrium we compare a theory, consisting of a few general principles, with a set of intuitions and we regard the theory with the best fit as most justified. But this assumes that the correct morality takes the form of a theory consisting of a few general principles, from which propositions regarding specific cases can be derived. This assumption is open to doubt, because moral knowledge may consist in possession of a skill rather than in grasping general principles and being able to derive propositions from them (1982: 329). More precisely, perhaps it consists in a different skill to this derivation skill and one that cannot be converted into propositional knowledge.

Note that David Wiggins also emphasizes practical knowledge which cannot be converted into general propositions, but when objecting to a different part of Rawls's philosophy, namely the conditions of the original position (2004: 485).

9. Objection from Andrei Marmor. To illustrate this objection, suppose that we are comparing two theories, T1 and T2, with our set of considered moral judgments and there are 30 of these, 30 intuitions. T1 fits with 29 intuitions whereas T2 fits with only 25 intuitions. So T1 is favoured over T2. In favouring it, we treat our intuitions as if they capture truth. T1 is regarded as more true than T2. But if this is how we should proceed, then we should regard any sacrifice of an intuition as a sacrifice of truth. The closest to truth one can get is the entire set of intuitions, even if no theory fits perfectly with all these intuitions. In which case, Rawls has no grounds for saying that it is sometimes acceptable to abandon an intuition to achieve a fit between intuitions and theory (Marmor 1991: 388-389).

Responding to this objection would seem to require arguing that the true moral theory takes the form of a system deduced from general principles; thus if the closest fit between a system and intuitions that we can get involves sacrificing an intuition then we are justified in doing so.

10. Objection from Hilliard Aronovitch. Reflective equilibrium allows for cheating (1996: 405). What follows is my elaboration of Aronovitch's thought. To grasp why one might say this, imagine that the only observations challenging Newton's physical theory are from the movement of Mercury. Still, the movement of Mercury is a problem. The Newtonian cannot deny this. In Rawls's reflective equilibrium, in contrast, there is nothing analogous. There is no way for some seemingly irregular thing to force one to acknowledge a problem for a theory. Should a single intuition not fit with a generally successful theory, then it is permissible to simply treat the intuition itself as false.

11. Another objection from Hilliard Aronovitch. The method is meant to be something liberals should accept, but actually they should not. We are supposed to enter reflective equilibrium open to potentially revising any commitment, but some commitments are non-negotiable for liberals, such as freedom of religion and the right to non-violent dissent (1996: 411).

12. Objection from Klemens Kappel. In reflective equilibrium, we try to achieve a theory which is systematic, simple, and general. If two theories are equal in other respects, such as how many intuitions they fit with, but one achieves these qualities to a greater degree, e.g. it consists of three rather than four principles, then we should prefer it. But what is the justification for aiming to achieve these qualities? There needs to be some justification, because it is open to doubt that a true moral theory will have them. However, Kappel argues that the prospects for justifying the aim to achieve these qualities are poor. He considers and rejects various efforts. For example, if it is said that this is a prima facie attractive way of engaging with moral theorizing, we will probably find that people differ about what is a prima facie attractive way in relation to morality (2006: 140-141).

13. Another objection from Peter Singer. Heuristics are rules of thumb that generally work well but which can also lead to errors. Cass Sunstein argues that there are such things as moral heuristics. For example, people's judgments on what treatment to choose can change depending on whether the same options are described in terms of lives saved or lives lost, because of general rules of thumb for processing such descriptions. Moral heuristics are pervasive enough that errors from them may affect considered moral judgments. Singer's response is that, given the findings from empirical research about pervasive moral heuristics, we should not compare moral theories with considered moral judgments (2005: 561). Note that Singer's thinking has been challenged by Joakim Sandberg and Niklas Juth (2011).

14. Objection from Timothy Williamson. Williamson poses the question of why we should regard reflective equilibrium as a description of a philosophical method at all. He concedes that philosophers make mutual adjustments of general theory and judgments about specific cases, but without a clarification of the nature of the evidence provided by the intuitions entered into the reflective equilibrium procedure – what kind of evidence is this? how can it function as evidence? – there is no reason to regard this as a method (2022: 246). Presumably, Williamson would regard a tea leaf reader as potentially specifying her method – "These are how the leaves reveal the future" – but he would regard it as a bad method based

on false answers; whereas with reflective equilibrium he thinks we have not even been provided with a description of a method owing to not addressing the questions.

15. Objection from T.H. Irwin. Suppose that a theory, T1, is subject to reflective equilibrium tests and equilibrium is achieved, at the expense of a few of our considered moral judgments. So now our considered moral judgments are different as a result of the process. Suppose that a few years later a new theory, T2, is proposed and it is subject to the reflective equilibrium process. But we will be asking whether it fits with our current set of considered moral judgments, the ones that emerged from the assessment of T1. It would be unsurprising if those judgments favour T1 over T2, since there has been some revision of our considered moral judgments in order to achieve a fit with T1.

It seems unfair to test T2 against these. T2 gets compared with intuitions adapted for T1 merely because T1 was historically first to be subject to reflective equilibrium. What Rawls needs is a fixed set of moral intuitions to compare theories with over time (2009: 901). Note: I cannot see how Rawlsians can accept that there is a sufficiently rich body of such intuitions to work with. In order for reflective equilibrium to be a non-foundationalist procedure, one has to enter it with an openness to the possibility of revising some of the intuitions with which a theory is compared.

16. Objection from Thomas Kelly and Sarah McGrath. Rawls asks us to set aside some moral judgments as unsuitable for relying on, rather than entering them into the reflective equilibrium method. I earlier mentioned judgments made while hesitant, upset, or frightened. He also refers to judgments which the judger has a strong interest in being true. But Kelly and McGrath ask us to consider the proposition "a person of color should not receive lesser consideration in virtue of being a person of color." A person of color has a strong interest in this proposition's being true, so are they not to enter it into the procedure, because they cannot reliably assess its truth, as Rawls suggests? Kelly and McGrath say that the person nevertheless has enough justification for the proposition to enter it (2010: 348-349).

17. Another objection from Thomas Kelly and Sarah McGrath. There is nothing in Rawls's procedure which prevents a person with evil moral intuitions, e.g. "In this situation I am morally required to kill a few random people," from achieving reflective equilibrium (2010: 347).

18. Objection from Danny Frederick. Reflective equilibrium is based on the ideal of a static condition in which everything is in harmony (the theory and the intuitions), but Frederick thinks that given the aim of improving our moral understanding, it is better to instead embrace an ideal of endless improvement (2016: 458). This will involve actively trying to uncover intuitions that do not fit with current theory, for instance in response to a novel thought experiment, trying to find inconsistencies, and trying to formulate new rival theories, rather than just waiting for them to occur to us (2016: 450). Disequilibrium helps advance knowledge. Note that Rawls's conception of the reflective equilibrium method is that it involves eliciting intuitions regarding possible situations, rather than merely settling for coherence between a theory and intuitions about the actual, and that equilibrium is then very hard to achieve (1999: 43).

19. Another objection from Danny Frederick. Frederick says that reflective equilibrium will allow for ad hoc ways of resolving inconsistencies, which are ways that teach us nothing. For example, if one is presented with an example of a situation in which one's theory

regarding the costs of religious fundamentalism does not apply, one can qualify the theory with "Except for that situation." Rawls's preference for theoretical simplicity puts pressure against ad hoc moves, but Frederick suggests that this is not enough – it still allows us to make exceptions whenever we are not aware of other solutions (2016: 447).

20. Tensions with structuralist Marxism? Sally Haslanger proposes that there is some difficulty in reconciling reflective equilibrium with Louis Althusser's structuralist Marxism (2020). I suspect continental structuralism's theory of meaning entails an objection. (This is different to what she has in mind but I think it needs to be registered.) If one almost reaches reflective equilibrium and decides to get rid of an isolated judgment which does not fit with one's theory, the aim is to keep the others. But the contents of these other judgments may well change with the removal, because, given structuralism, meaning is determined by place within a total system. For example, if I abandon the judgment that the government should provide schemes to reduce social isolation but keep the judgment that the government should prefer an economy which is better for the worst off group, my concept of a person may well change into something more economic, e.g. I can say that persons are social beings but to be that is, in normal conditions, to trade to realize a life plan. Since the worst-off group is defined as composed of persons, I am not keeping the same judgment. I can still check whether there is a fit with the new judgments but one worry is that was not my aim.

Also: I have made various objections, which I have not presented here, and defended the procedure against some above.

References
Aronovitch, H. 1996. Reflective Equilibrium or Evolving Tradition? *Inquiry* 39:399-419.
Brandt, R.B. 1979. *A Theory of the Good and the Right.* Oxford: Clarendon Press.
Brun, G. 2014. Reflective Equilibrium Without Intuitions. *Ethical Theory and Moral Practice* 17: 237-252.
Frederick, D. 2016. Defective Equilibrium. *Organon F* 23: 443-459.
Haslanger, S. 2021. Political Epistemology and Social Critique. In D. Sobel, P. Vallentyne, and S. Wall (eds.), *Oxford Studies in Political Philosophy*. Oxford: Oxford University Press.
Kappel, K. 2006. The Meta-Justification of Reflective Equilibrium. *Ethical Theory and Moral Practice* 9 (2): 131-147.
Kelly, T. and McGrath, S. 2010. Is Reflective Equilibrium Enough? *Philosophical Perspectives* 24: 325-359.
Irwin, T.H. 2009. *The Development of Ethics: A Historical and Critical Study*. *Volume III: From Kant to Rawls*. Oxford: Oxford University Press.
Marmor, A. 1991. Coherence, Holism and Interpretation: The Epistemic Foundations of Dworkin's Legal Theory. *Law and Philosophy* 10 (4): 383-412.
Rawls, J. 1999 (revised edition). *A Theory of Justice.* Cambridge, Massachusetts: Belknap Press.
Raz, J. 1982. The Claims of Reflective Equilibrium. *Inquiry* 25 (3): 307-330.
Sandberg, J. and Juth, N. 2011. Ethics and Intuitions: A Reply to Singer. *The Journal of Ethics* 15(3): 209-226.
Scanlon, T.M. 2003. Rawls on Justification. In S. Freeman (ed.), *The Cambridge Companion to Rawls.* Cambridge: Cambridge University Press.
Singer, P. 1974. Sidgwick and Reflective Equilibrium. *The Monist* 58 (3): 490-517.
Singer, P. 2005. Intuitions, heuristics, and utilitarianism. *Behavioral and Brain Sciences* 28 (4): 560-561.

T.R. Edward
10th July 2023

Tersman, F. 2018. Recent work on reflective equilibrium and method in ethics. *Philosophy Compass* 13: 1-10.

Wiggins, D. 2004. Neo-Aristotelian Reflections on Justice. *Mind* 113 (451): 477-512.

Williamson, T. 2022 (second edition). *The Philosophy of Philosophy*. Hoboken: John Wiley and Sons.