



<http://www.diva-portal.org>

Preprint

This is the submitted version of a paper published in *Philosophy Compass*.

Citation for the original published paper (version of record):

**Tersman, F. (2018)**

**Recent Work on Reflective Equilibrium and Method in Ethics**

*Philosophy Compass*, 13(6): e12493

<https://doi.org/10.1111/phc3.12493>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-337967>

This is a preliminary version of the article. It has been published in final form in *Philosophy Compass*, Wiley-Blackwell (2018, Early View, DOI: 10.1111/phc3.12493).

# Recent Work on Reflective Equilibrium and Method in Ethics

Folke Tersman

*Uppsala University*

## *1. Introduction*

If we wish to make up our minds about what to believe about some matter we may want to ensure that we reach a conclusion that is justified by the evidence and backed by good reasons. This holds not only for straightforwardly empirical issues, such as whether there is human-induced global warming, but also if the matter is moral and concerns whether a state of affairs is just or whether a decision is permissible.

It might seem less clear, however, what it *takes* for a moral belief to be justified. Philosophers commonly answer that question by invoking the idea of reflective equilibrium (hereafter IRE), where the phrase “reflective equilibrium” denotes a state we approach by resolving inconsistencies among our moral views and by revising them so that they provide mutual support to each other. IRE is a theory about the conditions under which moral beliefs are justified, but is also associated with a method for obtaining such beliefs.

IRE was first articulated by John Rawls (1951, 1971) and has since been the object of several waves of criticism. Still, it remains the most prominent approach to questions about justification and methodology in ethics. What accounts for its resilience? Tristram McPherson suggests (2015) that ethicists overlook the flaws of the approach, which he considers to be fatal, only because doing so allows them to continue to do what they are already doing

without getting “bogged down in methodological enquiry” (27). The aim of this paper, in contrast, is to argue that IRE actually merits its prominence. I shall primarily focus on recent discussions. Some recent objections are related to general arguments against the possibility of moral knowledge, such as so-called “evolutionary debunking arguments”.<sup>1</sup> It is held that those arguments not only second earlier worries about IRE but also raise new ones. Contrary to those suggestions, I argue that nothing that has come to light in those discussions decisively undermines IRE. If anything, recent discussions illustrate its strengths.

## *2. The idea of reflective equilibrium*

The method that is associated with IRE can be described as follows. In determining what to think about moral issues, we should begin by identifying which of our views about justice, rightness and other moral categories are most considered (our “considered moral judgments”). These are judgments that are held with confidence, not (or not evidently, anyway) distorted by self-interest and prejudice, and based on adequate non-moral information and sound inference patterns (see, e.g., Rawls 1951 and 1971, 20-1 and 47-8). They may include both general views (“All interests deserve equal consideration”) and judgments about particular cases (“The OJ Simpson trial was fair”). We are then to proceed by determining which of a set of competing moral principles best fits with our considered judgments. If there are nevertheless conflicts between that principle and our judgments we are to go back and forth between them and modify our views until coherence is achieved—a process in which no element is immune to revision (Rawls 1975, 8). IRE implies that if we reach that state, then our moral beliefs are justified. Their justification is, as Rawls puts it, “a matter of mutual support of many considerations, of everything fitting together into one coherent view” (1971, 21).

<sup>1</sup> See Katia Vavova’s Compass article about this topic (Vavova 2015).

The account of IRE and its method that I have just given is lacking in detail and raises legitimate questions. What does the relevant type of coherence require? What should guide us when we go back and forth? However, we do not need precision about all details to discuss the objections the paper focuses on, and what I am going to say by way of explicating IRE will be limited to the aspects relevant here.<sup>2</sup> I begin by making some general remarks.

First, I conceive of IRE as a theory about when moral beliefs are justified relative to the aim of uncovering the *truth* about moral issues (i.e., a theory about when they are “epistemically justified”).<sup>3</sup> This conception of IRE is not uncontroversial, and Rawls distances himself from it in some writings. He sometimes stresses that the point of the IRE method is not to discover moral truths but to facilitate a certain descriptive-seeming project, namely to get a better understanding of the moral sensibilities we in fact have (Rawls 1975). However, it is the interpretation according to which IRE concerns when moral beliefs are epistemically justified that makes best sense of the objections that I discuss in this paper.

Second, we reach a reflective equilibrium through achieving coherence among our moral beliefs. The coherence of a set of beliefs is a matter of the logical, explanatory and evidential relations that obtain among their contents. Coherence entails not only the absence of logical conflicts, but also that the members are connected in such a way that each obtains conditional support from the others. If we judge, for example, that it is wrong to do  $\phi$ , then we can increase the coherence of our set of beliefs by adopting a principle whose truth would explain the wrongness of  $\phi$ . On the interpretation that renders IRE the most controversial, such relations provide the *sole* source of the justification we obtain through reaching a reflective

<sup>2</sup> See Tersman 1993 for a discussion of some of the aspects, and Eggleston 2010 and Welch 2014 for two innovative recent elaborations of IRE.

<sup>3</sup> The view that moral views can be true is controversial. However, many different claims about the nature of moral thinking are consistent with it, including versions of the expressivist position that moral opinions are in fact conative attitudes.

equilibrium. There are versions of IRE that deny that coherence is sufficient and that require some of the input elements to have a kind of backing that is independent of coherence relations. However, I wish to explore the prospects of defending IRE without making such departures from its fundamentally coherentist character.<sup>4</sup>

Third, IRE implies that having reached a reflective equilibrium is sufficient for having justified moral beliefs, but is consistent with thinking that some of our moral beliefs have some degree of justification even before we reach that state. What is necessary for a moral belief to have any degree of justification, according to IRE, is that it is connected through the relevant relations to a significant subset of the rest of the subject's beliefs. But such a connectedness may obtain even if she has not managed to bring *all* her considered judgments and moral beliefs into equilibrium, as is the ultimate aim of the IRE method. To specify the level of connectedness that is necessary and sufficient for a given degree of justification will be hard. But that is a problem that has no bearing on what follows.

### *3. The early criticism*

Much of the early criticism against IRE focused on the fact that the outcome of the process Rawls envisages will depend on the starting points (the considered judgments). The early critics argued that if there is no independent reason to trust those starting points (garbage in), not much is achieved by bringing them into equilibrium (garbage out). (See, e.g., Brandt 1979, Hare 1975, Lyons 1975 and Singer 1974.)

For example, Peter Singer points out that the fact that our considered judgments are held with confidence and do not seem to be influenced by self-interest is consistent with their

<sup>4</sup> See Ebertz 1993 and Holmgren 1989 for discussions about whether IRE is best seen as a purely coherentist theory.

deriv[ing] from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past (1974, 516)

Singer's point is that if factors like these underlie our initial judgments, there is no reason to trust the conclusions we reach by taking them as starting points. Singer acknowledges that the method of reflective equilibrium allows the inquirer to revise some of her initial judgments, but he denies that this makes much difference. If the first move is to seek principles that match those judgments, then the revisionary element is in his view too limited.

An influential response to this complaint was to follow Norman Daniels and stress that the aim of the method is not just to reach a "narrow" equilibrium that only involves principles and moral judgments, but a "wide" one that also includes certain philosophical background beliefs, such as (say) a theory about the nature of persons, as well as scientific theories.<sup>5</sup> The point is that if we must also take those theories into account, then we will be required to make more extensive revisions among our initial judgments.

However, although the search for a wide equilibrium may lessen the influence of our initial judgments, it will not eradicate it altogether. It might therefore seem that Daniels' response does not get at the heart of the critics' worry. Moreover, there is a related objection which voices similar concerns but does not focus on the extent to which IRE requires revisions of our initial judgments.

The point of departure of the alternative argument is the idea that people have widely divergent considered judgments, due for example to differences in cultural belonging. By following the IRE method these people will end up with different and conflicting moral

<sup>5</sup> Daniels 1979. Michael DePaul has also made important contributions to this discussion (1993 & 1998).

beliefs. IRE implies that if we successfully follow the IRE method and reach a reflective equilibrium then the resulting moral beliefs are justified. However, it is plausible to hold that a method generates justified beliefs only if there is reason to think that a successful employment of that method increases one's chances of obtaining correct beliefs. And, given that people could follow the IRE method flawlessly and still end up with massively divergent beliefs, it is hard to see that there is such a reason in the case of that method.<sup>6</sup> In my view this objection is still the most troubling one, and I shall return to it towards the end.

#### *4. Evolutionary debunking and coherence*

Let us move on, however, to some more recent challenges. One such challenge is related to so-called "evolutionary debunking arguments" against the existence of moral knowledge. On an evolutionary account of morality, we have the moral convictions that we have (in part) because the possession of them prompts us to behave in ways that promote survival and procreation. Debunkers think that this account undermines the justification of our moral beliefs because it shows that they are unreliable, in the sense that the factors that cause them are entirely unrelated to their truth.<sup>7</sup>

The debunkers' challenge does not pertain just to IRE, but it figures nevertheless in an objection to IRE that McPherson offers,<sup>8</sup> McPherson suggests that a theory such as IRE is plausible only if it "vindicates" ethical inquiry by saying at least something distinctive and

<sup>6</sup> At least unless we adopt an implausible form of relativism according to which moral truth varies from person to person. See Haslett 1987, Little 1984, Singer 1974, and Stich 1988 for this objection. For a general discussion about how moral disagreement can generate skeptical implications, see Tersman 2006, esp. chapter 3 & 4.

<sup>7</sup> See, e.g., Street 2006, Joyce 2006a, Joyce 2006b, and Kahane 2011. See also Vavova 2015 for an exposition and discussion of the debunking strategy. It is often assumed that the skeptical problems it is meant to highlight only arises given a view about moral truth according to which it is relevantly independent of our beliefs and attitudes.

<sup>8</sup> For another critique of IRE that appeals to an evolutionary debunking argument, see Singer 2005. For a defense against Singer's objections, see Tersman 2008.

informative in response to skeptical challenges like the debunking one (2015, 655). He then argues that IRE fails to meet that constraint. *Pace* McPherson, I shall argue that IRE *does* give a distinctive and informative answer to debunking arguments, and thus succeeds in vindicating ethical inquiry in McPherson's sense.

IRE does this by facilitating a certain type of response to those arguments. That response urges that whether the Darwinian factors that have allegedly shaped our moral beliefs are unrelated to the truth of those beliefs depends on what the truth *is*. On some assumptions about the moral facts (about what makes actions right, etc), they may be unrelated. On others, however, they are not.

David Enoch (2010), for example, points to the idea that survival is (at least somewhat) good. The evolutionary account suggests that we are disposed to believe that actions are right to the extent that they promote (our) survival. If survival is good, then the account thus also suggests that we believe that actions are right to the extent that they promote something good. Now, if an action promotes something good this contributes to making it right, so we may conclude that the evolutionary account shows that our rightness beliefs are at least "somewhat in line" with the facts about rightness (430).<sup>9</sup>

There is a worry that the move Enoch makes here is illegitimate and begs the question against the debunkers.<sup>10</sup> After all, the claim that survival is good is a substantively moral one, and the assumption that we may justifiably rely on the truth of such a claim directly conflicts with the conclusion the debunkers seek (i.e., that all moral beliefs are unjustified). IRE contributes to the discussion by neutralizing this worry.

<sup>9</sup> For other versions of this strategy, see Brosnan 2011, Schafer 2010, Skarsaune 2011, and Wielenberg 2010.

<sup>10</sup> Vavova 2015 addresses this objection.

If coherence is the sole source of justification, as IRE insists, then information about the causes of our moral beliefs undermines their justification only if it decreases the coherence of our system. It may decrease the coherence, for example, through being in tension with some of the beliefs we initially held. Debunkers think that there is such a conflict between the evolutionary account and the conviction that our moral beliefs are reliable. They also think that we should restore coherence by dropping the latter conviction, as well as the beliefs themselves. However, if we would achieve an even greater level of coherence by adopting some claim P from which it follows that the evolutionary account does *not* exclude reliability of our moral beliefs, then that is the revision the IRE method recommends. Whether P is a moral claim is irrelevant. What matters is only how well P in turn coheres with the rest of our system. According to IRE, then, the soundness of a response to the debunking challenge is independent of whether it relies on a moral claim.

IRE addresses the debunking challenge by exploiting the fact that we can modify our moral beliefs as to absorb the evolutionary account into our system and still coherently deny the debunkers' thesis that our moral beliefs are unreliable. When faced with the possibility of such modifications, a debunker could change her challenge and try to reach the skeptical conclusion she pursues via some other thesis. She could grant that the evolutionary account does not by itself exclude the reliability of our moral beliefs and just make the point that it fails to *guarantee* their reliability. However, it is hard to see how that weaker claim gives rise to any urgent skeptical worries (see Vavova 2015 for further discussion).

##### *5. Moral intuitions and experimental philosophy*

McPherson also complains that IRE fails to help us resolve “the central controversies of moral methodology” (655). One of the controversies he has in mind concerns the role of moral intuitions. A moral intuition is an attitude toward some moral claim that we form simply by

considering its content and without consciously applying some principle. For example, when reflecting on a scenario where we can benefit five persons by sacrificing one, it may simply strike us as obvious that this would be wrong.

One question that could be raised about intuitions is whether attitudes of this kind provide evidence, in the following sense: if a principle conflicts with our intuitions, then that is a reason to think that principle incorrect. Although this may seem plausible, critics have highlighted certain empirical findings about the sources of our intuitions to undercut the assumption that intuitions provide evidence in this sense. As we saw with the evolutionary account, the critics' idea is to show that the target attitudes are influenced by factors that are irrelevant to their truth, such as the order in which scenarios are presented.<sup>11</sup>

Whether advocates of IRE are committed to the view that moral intuitions provide evidence is disputed. They *are* committed to thinking that considered judgments play that role, but it is a further question whether considered judgments are intuitions (see Brun 2014 for a good discussion). Nevertheless, unlike what McPherson suggests, IRE does help to resolve the controversy about intuitions. It does so by incorporating a distinctive view about when we are permitted to take *any* kind of attitudes to provide evidence, be it considered judgments, intuitions or something else.

My argument here turns on a particular way of elaborating IRE. I am assuming, crucially, that coherence of the kind IRE requires is not only a function of how our moral views are related to each other; it also depends on how our moral beliefs are integrated into our system of beliefs as a whole. Such integration is enhanced when we have second-order beliefs about a subset of our first-order moral views that indicate that those views are reliably formed through being generated by processes that ensure that many of them are correct. We do have such

<sup>11</sup> For some relevant studies, see Doris and Plakias 2008, Nichols, Stich, and Weinberg 2003, Nichols and Ulatowski 2007, and Swain, Alexander, and Weinberg 2008.

second-order beliefs in the case of our observational beliefs. What we know about how our sense organs work gives us reason to think that observational beliefs are usually correct when formed in favorable conditions. This defense helps to justify the evidential weight that is assigned to observations for example in the sciences. IRE says that our system needs to include resources for a similar defense of the evidential role of some of our *moral* views.<sup>12</sup>

IRE also imposes certain constraints on such a defense. While there is no bar to invoking substantive moral claims in that defense, in the way Enoch's proposal illustrates, those claims must however cohere with the rest of our moral beliefs. Moreover, our defense must be able to accommodate empirical findings of the kind mentioned above. Suppose, for example, that we invoke a certain factor (say, empathy) in order to justify that a certain moral intuition was reliably formed; but then we find that same factor operative in the formation of someone else's conflicting intuition. Then we must be able to explain away the dissent in a way that allows us to have continued confidence that the factor has had benign effects in our own case. In the absence of such an explanation, our defense is defective and we may have to subtract that judgment from the set we rely on.

IRE thereby provides a recipe for a resolution of the controversy about intuitions. We are entitled to think that moral intuitions provide evidence to the extent that we are able to offer a defense of their reliability that satisfies the constraints just mentioned. The idea that we need such a defense can also be accommodated by other positions, but IRE interprets the requirement in a distinctive way, for example by allowing that it may involve moral claims.

## 6. *Astrology and theology*

As just mentioned, I am interpreting IRE as holding that the justification of our moral views depends on how well they are entrenched in our total system of beliefs. This feature helps IRE

<sup>12</sup> The plausibility of this demand is acknowledged by Norman Daniels (1979, 271-2).

to address the worry that the conditions it takes to be sufficient for justification are implausibly weak (the “garbage in/garbage out” objection).

One way of arguing that the conditions are too weak is to point to other domains where it seems obvious that the conditions in question are insufficient. For example, suppose someone ended up with an internally coherent set of astrological beliefs by applying the IRE method. Surely that is not enough to make her beliefs justified. But if it is not enough in the case of astrological beliefs, it is also not enough in the case of moral ones.

Is the analogy convincing? Thomas Scanlon stresses that there is a crucial difference between ethics and astrology (2003 & 2014). While astrology involves causal claims about physics and psychology, ethics does not. The causal claims astrologers make are inconsistent with our best theories about those matters, which is why astrological beliefs are unjustified. Moral beliefs, by contrast, are consistent with the theories in question. An advocate of IRE can therefore agree that bringing one’s astrological beliefs into equilibrium does not make them justified, and yet still defensibly maintain that the IRE method generates justification in ethics.

Thomas Kelly and Sarah McGrath respond to Scanlon’s observation about astrology by pointing to another domain, namely “theology understood as ‘transcendent metaphysics’” (2010). To interpret theology in that way is to assume that religious claims “have absolutely no observable consequences or upshot for the empirical world” (343). If this is how theological claims should be construed, a person’s theological beliefs are just as sheltered from tensions with empirical theories as our moral beliefs are. Still, we would hesitate to judge the internal coherence of a set of theological beliefs sufficient for justification. The point is once again that IRE makes justification come too cheaply.

Kelly and McGrath are right to insist that the mere fact that our beliefs in an area are sheltered from tensions with outside beliefs does not make them justified, even supposing that

they are internally coherent. But an advocate of IRE can agree with that point, since IRE takes the justification of our moral beliefs to depend on how well they are entrenched in our total system. Being inconsistent with beliefs outside of the domain is just one way in which they can fail to be so entrenched. Another is through failing to obtain *support* from those other beliefs, as would be the case for theological beliefs if theology is understood as transcendent metaphysics. IRE can therefore explain why both astrological and theological beliefs, though internally coherent, are not justified.

Someone might object that this response saves IRE from setting the bar for justification too low only by setting it too high. For if—as Scanlon seems to think—ethics is just as insulated from tensions with empirical beliefs as transcendent metaphysics is, then moral beliefs would also fail to obtain support from such beliefs. However, Scanlon’s isolationist view about ethics is questionable, even granted that moral beliefs do not involve any causal claims. Empirical facts help to justify our moral beliefs in many ways. For example, if Enoch’s attempt to explain the reliability of our moral beliefs via an evolutionary account of their origins succeeds, then it illustrates how an empirical theory can play a crucial role in their justification. Of course, empirical facts may only be able to support our moral beliefs relative to a set of background assumptions which includes a moral claim that explains their relevance (such as, in Enoch’s case, the claim that survival is good). But that type of relativism is a general feature of evidential support and does not lessen the significance of the existence of such relations.

### *7. An arbitrary mix?*

IRE is often described as an “internalist” theory of justification, in that it implies that what we are justified in believing is determined by our subjective attitudes. Internalist theories are

typically contrasted with “externalist” ones that stress that a belief’s justification depends on factors beyond the subject’s attitudes and mental states, such as the causes of the belief.

IRE’s internalism is partly manifested in the fact that Rawls assumes that whether a judgment is a *considered* judgment (and therefore a legitimate starting point for ethical inquiry) is independent of its content. McPherson points out that this means that IRE does not rule out someone’s starting from considered judgments that display “moral idiocy” (663). Is that a problem? It is true that IRE permits us to rely on judgments that *others* think are idiotical. This is just a straightforward implication of its internalism. However, it is consistent with IRE to argue that if we learn that someone dismisses our considered judgments as being idiotical this gives us a reason to drop them, unless we can explain away her dismissal in a way that leaves the reliability of the judgment intact.

What I want to focus on in this section, however, is a different worry that McPherson also brings up in this connection, namely that IRE is not *purely* internalist but also involves externalist elements. As just noted, Rawls assumes that a moral judgment can be a considered judgment regardless of its content. Rawls also holds, however, that the subject has reached a reflective equilibrium only if her set of beliefs is logically consistent, independently of her own assessment of the significance of consistency. This may seem *ad hoc* and arbitrary, as there are controversial philosophical questions also about the nature and weight of consistency. (See Williamson 2007, 244-6.) Why is the subject’s attitude thought to be decisive in the first case and not the other?

In response, I would urge that complaints about arbitrariness should not be assessed in a vacuum. Whether a position is objectionably arbitrary depends on what it is supposed to achieve and what evidence it is responsive to. Advocates of IRE are likely to insist that the relevant evidence here includes judgments or intuitions about when beliefs are justified. And since we generally have both intuitions that point in an internalist direction and intuitions that

suggest an externalist conception, *any* viable theory of justification is likely to constitute a mix of both types of elements.

### *8. Divergent equilibria*

In conclusion, I submit that none of the recent objections surveyed here really strengthens the case against IRE. However, a complete assessment of the position also requires revisiting the earlier worries. I shall briefly discuss the concern I said earlier I find most troubling, namely the one invoking the possibility of different and indeed inconsistent reflective equilibria across different subjects.

This possibility commits advocates of IRE to thinking that conflicting moral beliefs, and thus also false moral beliefs, could be justified. That is hardly in itself a problem. Justification does not guarantee truth, and two persons can be justified in having conflicting beliefs through having access to different evidence. But Kelly and McGrath, who sympathize with this objection, still think the possibility of divergent equilibria constitutes a problem for IRE, since “a good method for investigating a given domain should lead rational inquirers who impeccably follow that method to converge in their views over time” (341). We can call this “the no-convergence objection”.

In my view, critics of IRE often exaggerate the extent to which actual inquirers would fail to converge if they all flawlessly carried out the IRE method. The critics seem to ignore, for example, that IRE acknowledges that (known) disagreement is a consideration that may command revisions of one’s beliefs. Recall that IRE entails that whether we are justified in taking our considered judgments to provide evidence depends on whether our system contains positive reasons to think that they are reliably formed. We may attribute their reliability, for example, to certain intellectual skills. However, if we find that those skills are also possessed by people who reject some of the judgments, then we need to reconsider our defense of their

reliability and possibly remove the controversial ones from those we rely on. This would leave us with fewer considered judgments that are contested, which increases our chances of arriving at conclusions that overlap with those that others reach.

Of course, the existence of merely *possible* inquirers who would reach radically different conclusions cannot be ruled out; but the relevance of *their* dissent is unclear. The relevance depends on why one might think a method does not generate justified beliefs unless it produces convergence. The root idea here is that although truth and justification may come apart, there must be some sort of non-accidental connection between them. Justification must in some sense “increase our chances” of discovering the truth. However, the precise nature of the relation between truth and justification is disputed, and some conceptions of the relation are friendlier to IRE than others. The no-convergence objection presupposes an externalist conception to the effect that a factor (such as coherence among our beliefs) does not make them justified unless it in *in fact* moves us closer to the truth. But we can also imagine internalist conceptions, on which what matters is instead whether the factor can be seen to increase our chances of getting at the truth from our *own* perspective. A complete defense of the no-convergence objection would require arguments that help us to exclude that view (for a more extensive discussion, see Tersman 1993, chapter 5).

What might serve as premises for such an argument? Epistemologists usually proceed by supporting their conflicting claims with intuitions. However, since both internalism and externalism have intuitive support, it might be worthwhile to try alternative strategies.

Here we come face to face with tricky meta-epistemological issues about how to vindicate theories about justification. On one view, accounts of justification are aimed at

capturing the (or a) concept that people actually have in mind when assessing beliefs. On another conception, such accounts are rather to be seen as proposals for how the concept *should* be defined to serve certain theoretical or practical purposes. For example, we might need a definition of justified belief in order to articulate when a person ought to be blamed for having unintentionally caused a bad outcome (on the assumption that her blameworthiness depends on whether she was justified in thinking that it would not obtain). Personally, I am sympathetic to the second approach, and an idea worth exploring is that if we specify a purpose in precise enough terms, a distinctive conception of the relation between truth and justification may fall out, which in turn may help us to assess the no-convergence objection. McPherson makes another suggestion, namely that when discussing method in ethics, we should focus more on meta-ethical questions about the nature of moral beliefs and moral thinking.<sup>13</sup> What seems clear at any rate is that progress will require broadening the set of strategies we use when trying to settle controversies about IRE.<sup>14</sup>

<sup>13</sup> McPherson 2015, 668-70. See also Kappel 2006 for a good discussion of the meta-justification of IRE.

<sup>14</sup> Thanks to an anonymous reviewer for helpful comments. Special thanks to the section editor, Sarah Stroud, whose many perceptive and apt suggestions have significantly improved the paper.

## References

- Brandt, R. (1979). *A Theory of the Good and the Right*, New York: Oxford University Press.
- Brosnan, K. (2011). Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?, *Biology and Philosophy* 26: 51–64.
- Brun, G. (2014). Reflective Equilibrium Without Intuitions?, *Ethical Theory & Moral Practice* 17, 237-52.
- Daniels, N. (1979). Wide Reflective Equilibrium and Theory Acceptance in Ethics, *Journal of Philosophy*, 76, 256–82.
- Doris, J., and Plakias, S. (2008). How to Argue About Disagreement: Evaluative Diversity and Moral Realism, in Sinnott-Armstrong, W. (ed), *Moral Psychology Volume 2, The Cognitive Science of Morality: Intuition and Diversity*, Cambridge, Mass.: MIT Press, 303-31.
- DePaul, M. (1993). *Balance and Refinement: Beyond Coherentism in Moral Inquiry*. London: Routledge.
- DePaul, M. (1998). Why Bother with Reflective Equilibrium, in DePaul, M., and Ramsey, W. (eds), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, Lanham, MD: Rowman & Littlefield.
- Ebertz, R. P. (1993). Is Reflective Equilibrium a Coherentist Model?, *Canadian Journal of Philosophy* 23, 193-215.
- Eggleston, B. (2010), Practical Equilibrium: A Way of Deciding What to Think about Morality, *Mind* 119, 549-84.
- Enoch, D. (2010). The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it, *Philosophical Studies* 148, 413–38.
- Hare, R. (1975). Rawls' Theory of Justice, in Daniels, N., (ed.), *Reading Rawls*, Oxford: Blackwell, 81-107.
- Haslett, D. W. (1987). What Is Wrong with Reflective Equilibria?, *Philosophical Quarterly* 37, 305-11.
- Holmgren, M. (1989). The Wide and Narrow of Reflective Equilibrium, *Canadian Journal of Philosophy* 19, 43-60.
- Joyce, R. (2006a). *The Evolution of Morality*, Cambridge, Mass.: MIT Press.
- Joyce, R. (2006b). Metaethics and the Empirical Sciences, *Philosophical Explorations* 9, 133-48.
- Kahane G. (2011). Evolutionary Debunking Arguments, *Nous* 45, 103-25.

- Kappel, K. (2006). The Meta-Justification of Reflective Equilibrium, *Ethical Theory and Moral Practice* 9, 131-47.
- Kelly, T., and McGrath, S. (2010). Is Reflective Equilibrium Enough?, *Philosophical Perspectives* 24, 325-59.
- Little, D. (1984). Reflective Equilibrium and Justification, *Southern Journal of Philosophy* 22, 337-88.
- Lyons, D. (1975). Nature and Soundness of the Contract and Coherence Arguments, in Daniels, N., (ed.), *Reading Rawls*, Oxford: Blackwell, 141-68.
- McPherson, T. (2015). The Methodological irrelevance of reflective equilibrium, in Daley, C., (ed.), *Palgrave Handbook of Philosophical Methodology*, 652-74.
- Nichols, S., Stich, S., and Weinberg, J. (2003). Metaskepticism: Meditations in Ethno-Epistemology, in Luper S., (ed.), *The Sceptics*, Burlington, VT: Ashgate.
- Nichols, S., and Ulatowski, J. (2007). Intuitions and Individual Differences: The Knobe Effect Revisited. *Mind & Language* 22, 346–365.
- Rawls, J. (1951). Outline of a Decision Procedure for Ethics, *Philosophical Review* 60, 177-97
- Rawls, J. (1971). *A Theory of Justice*, Cambridge, Mass.: Harvard University Press.
- Rawls, J. (1975). The Independence of Moral Theory, *Proceedings of the American Philosophical Association* 48, 5-22.
- Scanlon, T.M. (2003). Rawls on Justification, *The Cambridge Companion to Rawls* (ed. Freeman, S.), Cambridge: Cambridge University Press, 139-67.
- Scanlon, T.M. (2014), *Being Realistic about Reasons*, Oxford: Oxford University Press.
- Schafer, K. (2010). Evolution and Normative Scepticism, *Australasian Journal of Philosophy* 88: 471–488.
- Singer, P. (1974). Sidgwick and Reflective Equilibrium, *Monist* 58, 490-517.
- Singer, P. (2005), Ethics and Intuitions, *The Journal of Ethics* 9, 331-52.
- Skarsaune, K. (2011). Darwin and Moral Realism: Survival of the Iffiest, *Philosophical Studies* 152: 229–243.
- Stich, S. (1988). Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity, *Synthese*, 74, 391–413.
- Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value, *Philosophical Studies* 127, 109-166.
- Swain, S., Alexander, J. and Weinberg, J. (2008). The instability of philosophical intuitions: Running hot and cold on Truetemp, *Philosophy and Phenomenological Research* 76, 138-155.

- Tersman, F. (1993). *Reflective Equilibrium. An Essay in Moral Epistemology*, Stockholm: Almqvist & Wiksell.
- Tersman, F. (2006). *Moral Disagreement*, New York: Cambridge University Press.
- Tersman, F. (2008). The Reliability of Moral Intuitions: A Challenge from Neuroscience, *Australasian Journal of Philosophy* 86, 389-405.
- Vavova, K. (2015). Evolutionary Debunking of Moral Realism, *Philosophy Compass* 10: 104-116. doi: 10.1111/phc3.12194.
- Welch, J. (2014). *Moral Strata: Another Approach to Reflective Equilibrium*, Springer.
- Weinberg, J., Nichols, S., & Stich, S., (2001). Normativity and Epistemic Intuitions, *Philosophical Topics* 29, 429–460.
- Wielenberg, E. J. (2010). On the Evolutionary Debunking of Morality, *Ethics* 120, 441-64.
- Williamson, T. (2007). *The Philosophy of Philosophy*, Oxford: Blackwell.