INTRODUCTION

Reflective equilibrium: conception, formalization, application—introduction to the topical collection

Georg Brun[1] · **Gregor Betz**[2] · **Claus Beisbart**[1]

Did you ever submit a grant proposal to a funding agency? Then, you have likely encountered the request to specify your research method. Anecdotal evidence suggests that philosophers often address this unpopular request by mentioning reflective equilibrium (RE), the method proposed by Goodman (1983 [1954]) and baptized by John Rawls in his "A Theory of Justice" (1971). Appeal to RE has indeed become a standard move in ethics (see, e.g., Daniels, 1996; Swanton, 1992; van der Burg & van Willigenburg, 1998; DePaul, 2011; Mikhail, 2011; Beauchamp & Childress, 2013). The method has also been referred to in many other branches of philosophy, e.g., in methodological discussions about logic (e.g., Goodman, 1983; Resnik, 1985, 1996, 1997; Brun, 2014; Peregrin & Svoboda, 2017) and theories of rationality (e.g., Cohen, 1981; Stein, 1996). Some philosophers have gone as far as to argue that RE is unavoidable in ethics (Scanlon, 2003) or simply *the* philosophical method (Lewis, 1983, p. x; Keefe, 2000, ch. 2).

The popularity of RE indicates that its key idea resonates well with the inclinations of many philosophers: You start with your initial views or commitments on a theme and try to systematize them in terms of a theory or a few principles. Discrepancies between theory and commitments trigger a characteristic back and forth between the commitments and the theories, in which commitments and theories are adjusted to each other until an equilibrium state is reached.

✉ Georg Brun
Georg.Brun@unibe.ch

Gregor Betz
gregor.betz@kit.edu

Claus Beisbart
Claus.Beisbart@unibe.ch

1   Institute of Philosophy, University of Bern, Länggassstrasse 49a, Bern 3012, Switzerland

2   Institute of Philosophy, Karlsruhe Institute of Technology, Douglasstraße 24, 76131 Karlsruhe, Germany

⚐ Springer

Given the popularity of RE, it is no surprise that the method itself has become a topic of philosophical research. For instance, early on, Daniels tried to disentangle the method from the specific application in Rawls's 1971 classic. He also promoted the idea of a wide reflective equilibrium that examines theories not just in view of commitments but also using background theories (Daniels, 1979). Elgin (1993, 2017) developed a comprehensive account of the method, suggesting it as an account of epistemic justification in general. Given the inherently controversial character of philosophy, it is no wonder that critics of RE have emerged too. Singer (1974), for instance, objected in his review of Rawls's "Theory of Justice" that RE is unduly subjectivist. Brandt (1985) complained that RE "seems to amount to no more than a re-shuffling of one's initial prejudices" (Brandt, 1985:7). More recently, Kelly and McGrath (2009) argued that RE does not suffice as justification for views that can be evaluated regarding their truth, as realists have it.

In the last few years, the discussion about RE has gained additional momentum, in line with a general surge in the interest in philosophical methodology, as testified, for example, in the interest in meta-metaphysics and conceptual engineering. In particular, some researchers have started providing formal RE accounts (Yilmaz et al., 2017; Beisbart et al., 2021, Baumgaertner, 2023). Attempts steering in this direction are well motivated by the fact that many discussions about RE have remained inconclusive because they relied on rather vague descriptions of RE.

Against this background, we have recently organized a conference about RE. The conference was intended to commemorate the 50th anniversary of the emergence of reflective equilibrium under this name in Rawls's *A Theory of Justice* (but postponed due to the COVID-19 pandemic). The many interesting and novel contributions motivated us to edit this topical collection in *Synthese*, issuing a call for papers that mentioned three focus areas. We first encouraged philosophers to propose and clarify *conceptions* of RE. As indicated, the discussion about RE is unlikely to make progress as long as it remains unspecific about what RE actually is. Secondly, we invited philosophers to illuminate RE using *formalizations*. Finally, we solicited work on *applications* of RE, be it reconstructions of previous arguments in terms of RE, new applications of the method or investigations of the consequences that RE might have for philosophical inquiry.

The resulting collection consists of 12 papers. We will now introduce them, roughly in the order *conception– formalization– application*, although some papers may not be unequivocally assigned to one of the three terms.

The first bunch of papers focusing on the conceptions of RE opens with the contribution by Michael W. Schmidt. His paper reacts to the fact that many RE accounts exist in the literature. The paper thus proposes a minimal working definition of the method of RE, which Schmidt contrasts with purely negative characterizations of RE. His minimal definition consists of four conditions that are intended to be necessary and jointly sufficient. Minimalist foundationalism asks the agent to start with their current beliefs and with theories they take to be relevant. Minimalist fallibilism requires the agent to be prepared to revise each belief. Moderate holism includes the conditions that the net should be cast as widely as possible– clearly, Schmidt is interested in wide RE– and that the system of beliefs be considered as a whole. Finally, minimalist rationality requires agents to identify the most plausible system

of beliefs given possible constraints on time and resources. Schmidt argues that other requirements, e.g. the demand to start from considered judgments only, should not be part of a minimal definition of the method of RE. Rather, such requirements may be justified on the basis of minimal RE, at least for certain applications. In the spirit of minimalist fallibilism, Schmidt is open to the possibility that his minimal definition might be improved.

Schmidt's minimal notion of RE may be a useful basis for discussing the commonalities between various accounts of RE. However, this leaves the possibility that some accounts of RE are more attractive than others. In his contribution, Wibren van der Burg argues that the conceptions of RE that Rawls himself developed suffer from the fact that they draw on two sources, viz. pragmatism and Kantian views. As a reaction, van der Burg develops a thoroughly pragmatist RE that is, so to speak, purged from the Kantian elements. In his project, he is not interested in the general idea that various elements should be brought into coherence, but rather takes Rawls's specific version of RE in the *Theory of Justice* as a point of departure to develop a method for moral philosophy. In his pragmatist spirit, van der Burg proposes that RE has three interconnected aims: right action, moral understanding, and self-improvement. He then discusses key aspects of RE. Regarding the input to RE, he pleas against restricting it to convictions felt with high confidence, and against excluding comprehensive views. In van der Burg's pragmatist understanding, coherence has three facets, viz. consistency, mutual support, and comprehensiveness. He further stresses that the agent should confront their view with various philosophical perspectives. In the final part of his paper, van der Burg defends the resulting conception of RE. He first argues that it is no problem that RE cannot be shown to reach truth or certainty since our views are bound to remain erroneous in many respects anyway. Nor is it a problem that pragmatist RE is not a strict method, or so he argues.

The next paper, by Claus Beisbart and Georg Brun, is concerned with the conception of RE in a more indirect way, starting from popular objections against RE. Although proponents of RE have developed replies to individual objections, there is a danger that the indeterminacy of RE permits to add a certain twist to the conception of RE when defending it against one objection while steering in a different direction when addressing another. The paper thus asks whether there is one version of RE that can be defended against all objections. To answer this question, Beisbart and Brun systematize the most prominent objections against RE. After the general objection that accounts of RE are not sufficiently informative to characterize anything of interest, they consider more specific objections, e.g. that RE draws on problematic input, that it produces garbage, if fed with garbage, and that it is too conservative. Beisbart and Brun also consider the objections that RE is tied to a problematic coherentist view of justification and that it is too demanding in practice. The strategy of the paper is first to articulate and clarify each objection, and then evaluate the main strategies that proponents of RE have used to rebut the objections. Ultimately, Beisbart and Brun argue that there is space for a conception of RE that can meet all the objections. This conception, however, may surprise both friends and foes of RE, since the authors propose, e.g., to decouple RE from coherentism.

Although the paper by Manuel Cordes is not aimed at a full account of RE, it is meant to improve the understanding of RE. Cordes focuses on a specific element that

is needed for (moral) RE, but seldom discussed, viz. belief about empirical facts. He argues that various matching relations that contribute to coherence depend on empirical beliefs. For instance, if a principle matches a moral judgment about a particular case, then this judgment needs to be derivable from the principle and suitable empirical beliefs about the case. Likewise, matching between two principles requires empirical beliefs. An immediate consequence is that changes in empirical beliefs can necessitate changes in moral principles. To illustrate this, Cordes refers to British law and the Double Jeopardy Principle, a principle that bars the reexamination of cases. This principle was given up due to developments in our empirical beliefs: It came to clash with spontaneous convictions that people happened to have in view of the method of DNA testing. Cordes argues that the sensitivity of RE to empirical information is a plus. However, his point that the matching relation depends on empirical beliefs also raises the question of whether an appeal to coherence can lead to a revision of empirical beliefs. Using an argument by van Inwagen as an example, Cordes argues that this is possible: Considerations of coherence can lead an agent to adopt the belief that determinism is false if the available empirical evidence does not favor determinism over indeterminism and if belief in indeterminism helps our moral beliefs to better match to each other.

The next bunch of papers uses formal models to study RE. In the first of these papers Finnur Dellsén chooses Bayesian epistemology as a formal framework to explicate RE. His focus is on equilibrium states rather than the process of equilibration. According to Dellsén, an epistemic agent is in a state of reflective equilibrium if their credences about a topic are probabilistically coherent and if every proposition she accepts follows from a comprehensive theory that is maximally confirmed by her credences. In this context, a theory is comprehensive just in case it entails answers to all salient questions in a context. Building on previous work (Dellsén, 2021), Dellsén argues that such comprehensive theories, and hence the accepted propositions in a state of RE, are deductively consistent and potentially closed. As advantages of this formalization of RE, Dellsén highlights that it fits tightly with Bayesian epistemology and accounts for the holistic character of RE. Note, too, that the model of an equilibrium state presented by Dellsén constrains accounts of the method of reflective equilibrium: To be effective, any such method has to allow agents to (gradually) revise incoherent initial credences and potentially inconsistent accepted propositions into an RE state. Explicating such a method would be a worthwhile endeavor that arises from Dellsén's paper.

While Dellsén focuses on formalizing the state of reflective equilibrium, Bert Baumgaertner and Charles Lassiter present a formal model of the step-wise process of equilibration in order to study the prospects of interpersonal convergence in RE. More specifically, they study the hypothesis that even minimal interpersonal influence may strongly increase the chance of consensus in a group of agents executing RE. Baumgaertner and Lassiter thus connect the scholarship on RE with social and network epistemology (e.g. Zollman, 2013). The formal framework that the authors use to explicate an RE process bears similarities to training a simple classifier model in machine learning: The model (principle/rule) agents are supposed to learn consists of a single center point, and classifies any example point as accepted iff its distance to the center is below a given threshold. In an individual RE process, agents

are presented with pre-labeled examples (intuitions), may revise their classification rule (i.e., shift the center point), and ultimately classify the example (possibly re-labeling it). In a collective RE process, agents communicate any classification they make within the network, thus setting a "precedent" for others. Given the precedents, agents are not only considering the initial label and the current model when classifying an example; rather, they also take into account how other agents have classified this example before, especially to break ties between the current model classification and the initial label. Baumgaertner and Lassiter use computer simulations to obtain the following results: More communication (denser networks) means that a group of RE agents is more likely to converge on the same principle. The computational study thus provides promising evidence that interpersonal convergence is feasible within RE.

The next two papers also study consensus formation using a formal model of RE. Richard Lohse investigates consensus formation in a setting that is motivated by Rawls's "Political Liberalism" (Rawls, 2005 [1993]). As is well–known, in this work, Rawls explores the possibility of an overlapping consensus on a political conception of justice. This is to say that citizens who hold different comprehensive doctrines agree on a conception of justice. However, Rawls did not consider how such an overlapping consensus might arise. Lohse thus proposes to study the conditions under which citizens who apply RE come to converge on a shared conception of justice. For this purpose, he uses a variant of the RE model proposed by Beisbart et al. (2021). This variant is more realistic because it assumes that citizens have bounded rationality and therefore cannot consider all possible theories in their attempt to systematize the commitments; they rather must proceed in a piecemeal manner. To model the specific setting that Rawls had in mind, Lohse defines a shared logical space of arguments connecting alternative ethical doctrines, a political conception of justice, and multiple further ethical and political statements. In this space, Lohse simulates parallel and pairwise independent RE processes that start from different initial commitments. By varying essential features of this space and observing whether the political conception of justice ends up as part of multiple final RE states, Lohse is able to discern how structural properties of the logical debate space affect the chance of overlapping consensus on a political conception. He finds that an overlapping consensus becomes more likely if more doctrines imply the conception of justice. This confirms Rawls's own suspicion, which is that an overlapping consensus on a political conception requires that most comprehensive moral doctrines actually imply the political conception. However, Lohse finds indications that a different scenario is friendly to overlapping consensus too.

Andreas Freivogel's paper studies a prominent objection against RE, namely the worry that RE is too weak as a method of justification because it does not sufficiently promote consensus formation. This worry has often been expressed but never been subject to a rigorous analysis. Freivogel uses the model by Beisbart et al. (2021) to investigate whether the worry is well-founded. For this purpose, he distinguishes between three versions of the objection. According to the first version, the equilibration process fails to attain a unique outcome in too many cases. Following Kelly and McGrath (2010), Freivogel further distinguishes between intrapersonal and interpersonal convergence on a unique outcome. Intrapersonal convergence is an issue

because the rules that govern the RE process leave certain choices open and thus at the discretion of one and the same agent. Using a big simulation study, Freivogel finds that intrapersonal convergence is fairly common, but not ubiquitous. Interpersonal convergence under which agents with different initial views end up with the same view is much rarer. Freivogel does not find this problematic but suggests that justification is, to some extent, pluralist. In its second version, the worry is that agents do not come closer to each other when they apply RE; in the worst case, they end up with incompatible positions. Freivogel investigates the plausibility of this worry by comparing random pairs of initial views with which agents start and the final RE state they reach. He finds that incompatibility is much more common in the initial stage than in the final stage. This means that applying RE makes it more likely that the agents come to hold compatible positions. Using a measure in which the distance between views can be compared, Freivogel finds further that the positions of agents get closer as the agents apply RE. The final version of the worry holds that RE allows for „anything goes": Every position can be justified using RE. To defuse this worry, Freivogel shows that views that form a RE are much rarer than positions that may be taken initially. The final conclusion is that the objection from lack of convergence is not sound because RE does promote consensus formation. However, there are cases in which RE fails to produce a unique result, even if only one agent is involved. This is not a problem if several positions may be justified to the same degree.

Folke Tersman takes a different route to defuse the worry that RE does not lead to consensus and may even give rise to radically divergent views. Since his paper does not draw on formal models, it marks the transition to the third bunch of papers, which focus on applications and consequences of RE. In the terms of Kelly & McGrath, Tersman concentrates on interpersonal convergence, focusing on moral views. His main idea is to draw on the philosophical debate about peer disagreement. In this debate, various authors have argued for a conciliatory attitude: If an agent observes that a peer disagrees with them on a proposition, they should consider this and, e.g., give both views equal weight. Tersman adopts this conciliationist position and argues that agents who disagree with peers on p, cannot take their judgment that p to be a considered judgment. Consequently, they should not use this judgment as an input for the method of RE. It then seems less likely that the agents come up with radically different equilibrium positions. Still, as Tersman acknowledges, there is a downside: Since the set of judgments that are used as a basis for RE is effectively shrunk if judgments are removed due to peer disagreement, it becomes more likely that no unique RE state will be reached because a smaller basis of judgments raises the chances of underdetermined choices. Tersman admits this, but rejects other worries about his argument. One possible objection to his argument starts from the view that a conciliatory attitude is only justified if there is positive evidence that the disagreeing parties are peers. In examples where there is broad disagreement with the others, this positive evidence is difficult to obtain. Tersman replies that we do agree on some moral principles and that, even in the face of disagreement, there are ways of ascertaining peerhood. Another objection is that, in the debate on peer disagreement, conciliationist positions are sometimes rejected in favor of the steadfast view. Tersman's reply is that the steadfast view is only plausible in cases that are dissimilar from his use case, i.e., the attempt to justify one's moral views. Note, in any case, that Tersman's

strategy is complementary to Freivogel's. While Freivogel stresses the power of the RE process, in particular the application of theoretical virtues, independently of the input to the method, Tersman focuses on the input, arguing that it may, in some cases, need revision.

Ben Martin discusses RE in logic and, thus, the context in which it was originally introduced. His main aim is to compare RE with his preferred epistemology of logic, "logical predictivism." Such a comparison faces the problem that many distinct descriptions of RE are available and most of them are very unspecific or intended for an application to ethics, not logic. Martin therefore takes Goodman's original description as a starting point and discusses how RE should be specified for application to logic. He argues that RE should (i) take judgements about the acceptability of individual inferences as input and (ii) deal with logical theories in a comprehensive sense comprising not only rules of valid inference but also, e.g., definitions, metatheory and a translation manual for natural language arguments; additionally, RE should (iii) include background theories (as demanded by "wide" RE) and (iv) account for the dynamics of agreement and competition in logical theorizing.

Martin's favoured alternative epistemology of logic, "predictivism", also exhibits features (i)-(iv), but emphasizes that choices between rival theories should be made based on a comparison of their explanatory power and predictive success. "Prediction" in this context means to use a logical theory for deriving results that can be tested against judgements. Martin explains why he thinks that, despite strong parallels, predictivism contradicts RE, as he interprets it, in three respects and why predictivism should be preferred: First, whereas RE takes no judgement to be in principle immune to revision, predictivism holds that some judgements are not up for revision, e.g. judgements about the acceptability of certain inferences in mathematical proofs. Second, whereas RE requires coherence only, predictivism additionally aims at explanation. Third, while RE holds that an agreement between judgements and logical theory is necessary for them to be justified, predictivism holds that it can be rational to endorse a theory that is inconsistent with certain judgements (or background theories) without triggering the need to change the theory or the recalcitrant judgements. For defenders of RE, Martin's argument raises interesting questions: Should they reject the features of predictivism or rather argue that RE does, or at least can, instantiate them as well.

Akira Inoue, Kazumi Shimizu, Daisuke Udagawa, and Yoshiki Wakamatsu propose how to apply RE in political philosophy. The topic of their case study, distributive justice, is very Rawlsian. Inoue et al. argue that the discussion about related positions, e.g. prioritarianism and sufficientarianism, suffers from two problems. First, philosophers appeal to their own intuitions rather than considering what the broader population thinks. Second, they rely on thought experiments, many of which are not relevantly similar to the situations to which the principles are intended to apply. But, as Inoue et al. argue further, it is impossible to decide a priori which thought experiments are legitimate. They propose to solve both problems using what they take to be an application of RE. They first present thought experiments to participants from the broader population. The resulting data are then evaluated using the so-called Akaike Information Criterion. Very roughly, the data are fitted to a model that is also supposed to be simple. Here, the quests for agreement with the data and for simplicity

are in line with RE's demands for agreement with pre-theoretic commitments and systematicity, respectively. Further, the fairly high weight put on simplicity means that not each reaction to an experiment is taken seriously. In other words, certain „deviant" data points are effectively not fitted using the model. This may indicate that a certain thought experiment was not legitimate or that some intuitions should be discarded. Inoue et al. apply this method to real data to show that a refined version of sufficientarianism does not beat prioritarianism.

Apart from Elgin's pioneering work (1996, 2017), explicit applications of RE to aesthetics have remained exceedingly scarce. Murray Smith's paper explores and substantiates the hypothesis that the formation of aesthetic experience and judgment does, and should, exhibit the structure of a reflective equilibrium process, even if it cannot be reduced to such a process. Relying on case studies and examples, Smith first explains how an understanding of a work of art can be reconstructed as the result of a process of mutually adjusting the appreciation of the specific work and its features on the one hand and artistic categories and theories on the other hand. Exploring parallels between aesthetics and ethics, he emphasizes that aesthetic interpretation, and in particular criticism, seeks reflective equilibria that are both collectively shared and wide. This means that the search for a reflective equilibrium needs to involve descriptive and evaluative judgments by other people, as well as more abstract background theories, especially in the case of artworks with a contested aesthetic status. Smith then defends his proposal against the two charges. According to the first objection, he over-intellectualizes the engagement with art. Under the second objection, he implausibly assumes that art, aesthetics, and aesthetic experience form a single topic that allows for a systematic account. Smith's reactions to the objections show that reflective equilibrium in aesthetics cannot only involve beliefs but also needs to include the content of other mental attitudes, which are more emotional or perceptual and less conscious.

All in all, the papers in this collection show the importance of RE as a subject for philosophical inquiry. In particular, formal accounts of RE can advance the discussion about various objections against it. As is common with formal work, such accounts raise new questions that open novel directions of research. On top of this, improved conceptualizations and formalizations of RE allow us to apply it more rigorously to various philosophical debates. We thus hope and think that RE will continue to draw the attention of philosophers also in the next 50 years. In the meanwhile, you are well advised to mention RE in your grant proposal.

References

Baumgaertner, B., & Lassiter, C. (2023). Convergence and shared reflective equilibrium. *Ergo, 10*, 673–705. https://doi.org/10.3998/ergo.4654

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics* (7th ed.). Oxford University Press. https://doi.org/10.1016/S0035-9203(02)90265-8

Beisbart, C., Betz, G., & Brun, G. (2021). Making reflective equilibrium precise. A formal model. *Ergo, 8*(15), 441–472. https://doi.org/10.3998/ergo.1152

Brandt, R. B. (1985). The concept of rational belief. *The Monist, 68*, 3–23.

Brun, G. (2014). Reconstructing arguments: Formalization and reflective equilibrium. *Logical Analysis and History of Philosophy, 17*, 94–129. https://doi.org/10.30965/26664275-01701006

Cohen, J. L. (1981). Can human rationality be experimentally demonstrated? *The Behavioral and Brain Sciences, 4*, 317–331. https://doi.org/10.1017/S0140525X00009092

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy, 76*, 256–282. https://doi.org/10.2307/2025881

Daniels, N. (1996). *Justice and justification. Reflective equilibrium in theory and practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511624988

Dellsén, F. (2021). Understanding scientific progress. The noetic account. *Synthese, 199*, 11249–11278. https://doi.org/10.1007/s11229-021-03289-z

DePaul, M. R. (2011). Methodological issues. Reflective equilibrium. In C. Miller (Ed.). *The continuum companion to ethics* (pp. lxxv–cv). Continuum. https://doi.org/10.5040/9781350217911.0007

Elgin, C. Z. (1996). Considered judgment. *Princeton University Press*. https://doi.org/10.2307/2653830

Elgin, C. Z. (2017). True enough. *MIT Press*. https://doi.org/10.1111/j.1533-6077.2004.00023.x

Goodman, N. (1983) [1954]. *Fact, fiction, and forecast* (4th ed.). Harvard University Press.

Keefe, R. (2000). *Theories of vagueness*. Cambridge University Press.

Kelly, T., & McGrath, S. (2010). Is reflective equilibrium enough? *Philosophical Perspectives, 24*, 325–359. https://doi.org/10.1111/j.1520-8583.2010.00195.x

Lewis, D. (1983). *Philosophical papers, Vol I*. Oxford University Press. https://doi.org/10.1093/0195032047.001.0001

Mikhail, J. (2011). *Elements of moral cognition. Rawls' linguistic analogy and the cognitive science of moral and legal judgement*. Cambridge University Press. https://doi.org/10.1017/CBO9780511780578

Peregrin, J., & Svoboda, V. (2017). *Reflective equilibrium and the principles of logical analysis*. Routledge. https://doi.org/10.4324/9781315453934

Rawls, J. (1999) [1971]. *A theory of justice. Revised edition*. Belknap Press.

Rawls, J. (2005) [1993]. *Political liberalism. Expanded ed.* Columbia University Press.

Resnik, M. D. (1985). Logic: Normative or descriptive? The ethics of belief or a branch of psychology? *Philosophy of Science, 52*, 221–238. https://doi.org/10.1086/289241

Resnik, M. D. (1996). Ought there to be but one logic? In B. J. Copeland (Ed.), *Logic and reality: Essays on the legacy of Arthur Prior* (pp. 489–517). Oxford University Press.

Resnik, M. D. (1997). *Mathematics as a science of patterns*. Clarendon Press.

Scanlon, T. M. (2003). Rawls on justification. In S. Freeman (Ed.), *The Cambridge Companion to Rawls* (pp. 139–167). Cambridge University Press. https://doi.org/10.1017/CCOL0521651670.004

Singer, P. (1974). Sidgwick and reflective equilibrium. *The Monist, 58*, 490–517. https://doi.org/10.5840/monist197458330

Stein, E. (1996). *Without good reason. The rationality debate in philosophy and cognitive science*. Clarendon Press.

Swanton, C. (1992). *Freedom. A coherence theory*. Hackett.

van der Burg, W., & van Willigenburg T. (Eds). (1998). *Reflective equilibrium. Essays in honour of Robert Heeger.* Kluwer. https://doi.org/10.1007/PL00000007

Yilmaz, L., Franco-Watkins, A., & Kroecker, T. S. (2017). Computational models of ethical decision-making. A coherence-driven reflective equilibrium model. *Cognitive Systems Research, 46*, 61–74. https://doi.org/10.1016/j.cogsys.2017.02.005

Zollman, K. J. S. (2013). Network epistemology: Communication in epistemic communities. *Philosophy Compass, 8*(1), 15–27. https://doi.org/10.1111/j.1747-9991.2012.00534.x