

Reflective equilibrium 51 years after *A Theory of Justice*

University of Bern, Switzerland, 12.–14. May 2022

Abstracts

Convergence and Shared Reflective Equilibrium

Bert Baumgaertner (Idaho) and Charles Lassiter (Gonzaga University)

We build a formal model of the reflective equilibrium method to better understand under what conditions a community of agents would achieve a shared equilibrium. We find that, despite guaranteeing that agents individually reach equilibrium and numerous constraints on how agents deliberate, it is surprisingly difficult for a community to converge on a small number of interpersonal equilibria. Agents tend to diverge from one another and extensive support is needed to limit divergence. This support can come in various forms, such as shared intuitions (considered judgments) about cases, starting with the same principles, updating rules in the same way, agreement on necessary conditions, and using similar strategies for handling tensions between a principle and intuitions about a case. Yet agent can still diverge. The problem is that in token applications of resolving a tension between a principle and intuitions about a case, agents have choice points on which they may not coordinate: ignore the intuition, ignore the case, change the principle, or change the tolerance for satisfying the principle. We believe the literature has underestimated this challenge of coordinating intrapersonal convergence and interpersonal convergence, and that convergence likely requires additional forces facilitated by interactions between agents.

The Dissolution of Comprehensive Moral Doctrines in Justice as Fairness

Paige Benton (Pretoria)

The focus of my paper is on demonstrating how and why reflective equilibrium poses serious risks to moral comprehensive doctrines in Rawlsian theory. I claim that the outcome of the method of reflective equilibrium requires the reshaping of comprehensive doctrines to align with the core values of a liberal constitutional democracy, identified via the process of reflective equilibrium. The significance of this alignment, I suggest, is the possible gradual dissolution of comprehensive moral theories that cannot fit in with the values of justice, and the categorical necessity to endorse liberal democratic norms of justice. This results in comprehensive moral doctrines in a society regulated by justice as fairness being reconfigured to specifically liberal democratic versions of their doctrines. I thus, argue that reflective equilibrium does not only help foster agreement on the core values of justice all persons share, but also requires comprehensive moral doctrines to be reshaped in light of the core values that form the basis for the overlapping consensus. The implication of this, is that reflective equilibrium helps to identify and promote a moral community in a Rawlsian society.

Reflective Equilibrium in Fair Allocation

Eleonora Cresto (Buenos Aires)

This paper has two main goals. First, I will discuss the extent to which agents rely on some sort of Reflective Equilibrium (RE) to deal with problems of fair allocation, such as cake-cutting. Cake-cutting problems fall within the domain of problems of distributive justice, broadly understood, so the question of the role of RE seems fitting. To the best of my knowledge, no one has addressed this topic before. I will argue that searching for a RE plays an important role in fair division, although the adjusting mechanism is normally hidden. The importance of the adjusting mechanism becomes clear once a number of crucial (and often neglected) properties of fairness become salient; for this I draw on previous work on the subject. Second, as a result of the previous discussion, I will be able to make a number of general conceptual observations on the structure and scope of RE.

Probabilifying Reflective Equilibrium

Finnur Dellsén (Reykjavik/Lillehammer)

This paper aims to explicate the celebrated notion of *reflective equilibrium* in a probabilistic framework for epistemic rationality. On the account developed here, reflective equilibrium consists in bringing into rational harmony two distinct kinds of commitments: credences and acceptances. Whereas credences are graded commitments that guide action, acceptances are outright commitments that jointly constitute one's overall intellectual worldview. Assuming that an agent is probabilistically coherent, these commitments are brought into rational harmony precisely when the agent accepts all and only propositions that follow from a maximally informative theory that is at least as probable as any other such theory. Put differently, the agent's overall worldview must be the most probable of all the alternative worldviews on offer. It can be shown that when an agent is in reflective equilibrium in this sense, the propositions accepted by the agent are bound to be logically consistent and closed under logical consequence. Moreover, recent formal work on abductive reasoning suggest that these propositions will be explanatorily virtuous. Thus the current account vindicates the idea that reflective equilibrium is closely tied to logical and explanatory coherence, even though the account itself appeals only to probabilistic coherence.

Dynamic Epistemic Equilibrium

Catherine Z. Elgin (Harvard)

Epistemic agents are finite and fallible. Our range is limited and some of what we accept is, no doubt, flawed. To achieve our epistemic and practical objectives, we employ the method of equilibrium to devise methods and practices that foster correction, refinement, and expansion of our current epistemic commitments. Traditional epistemology maintains that being in reflective equilibrium is at best indicative of acceptability. The criterion of success is non-fortuitously justified true belief. I argue that being in reflective equilibrium is constitutive, not merely indicative, of epistemic acceptability. A network of cognitive commitments in reflective equilibrium is as reasonable as any available alternative in the epistemic circumstances. That does not make it perfect or permanently acceptable. Such a network is susceptible of and probably in need of improvement. But it provides a platform for improvement. I argue that it should be designed to foster, not merely admit of, further gains. That requires that it enable critical reflection of its own ends and means. Integral to a dynamic understanding of a topic then is understanding our understanding of that topic. This enables us to recognize opportunities and obstacles to improvement.

Addressing the No-Convergence Objection Against Reflective Equilibrium With a Formal Model

Andreas Freivogel (Bern)

The no-convergence objection against reflective equilibrium (RE) is a prominent and longstanding line of criticism of RE. Roughly, it claims that RE is too weak as a method of justification because it fails to let views converge. Divergence among equilibria is suspected to arise from drastically different starting points and the method's lack of revisionary power, or from path-dependent adjustments during equilibration processes.

To overcome vagueness that besets the general discussion about RE, I rely on a formal model and operationalize the no-convergence objection in its framework. This includes measurable features that can track convergence as well as a merely streamlining baseline procedure for comparison. A computer implementation of the formal model allows to explore large ensembles of small RE simulations.

Findings suggest that the formal model can escape the no-convergence objection for a plausible range of parameter configurations. In addition, we learn from these configurations that a frequently overlooked aspect of RE should receive more attention in the informal debate: Theoretical virtues that are deployed in RE can strengthen the method.

The Applications of Wide Reflective Equilibrium. A Systematized Search and Critical Review

Kevin Helms (Bremen)

The wide reflective equilibrium (WRE) is intensively discussed in the scientific community, neglecting to look at its applications to date. The objective is to fill this gap by providing a critical overview of the published applications. Studies were searched that deal with WRE and are written in German or

English. All topics, disciplines and any publication forms were included. Key elements such as purpose, topic, categories and elements of levels as well as the adjustment process and used quality criteria were extracted.

The search returned 1,948 hits. Duplicates were removed. 537 titles and abstracts were screened. 22 of 199 studies met inclusion criteria for full texts. 38 applications were identified, of which 19 reached equilibrium. The applications themselves varied widely in terms of approach and the scope of elements considered. In particular, the adaptation process and quality criteria were rarely considered and explained transparently. On the one hand, the results show that the WRE is partly successfully applied but on the other hand, the number of applications is still very low. Therefore, there is a need for further research, especially with regard to the specification of the process steps with a focus on quality assurance.

Are Rawlsian Considered Judgements Analogous to Observation Reports?

Loren King (Waterloo)

John Rawls famously thought we could defend moral principles not by deducing them from self-evident moral axioms, or intuiting them, or by inferring them from non-moral facts. Instead, for Rawls as for several of his contemporaries, justification is a kind of balance between well-established moral commitments - considered judgements - and the systems of moral principles we build to clarify and organize those commitments. We seek a reflective equilibrium between our considered judgements and our principles, against a settled background understanding of the world, and that equilibrium is our justification.

Critics suspect we cannot have it both ways: either our principles need some foundation, or our considered judgements must be grounded in some way; these things cannot justify one another simply by being "in equilibrium". If we justify our principles in terms of their success in clarifying and organizing our considered judgements, then we must provide some account of why those judgements are morally privileged.

Roughly parallel to these debates, philosophers of science have been wrestling with the messy and reciprocal relationship between theory and experiment evident from studying how scientists actually ply their varied trades. While we often tell the story of scientific progress as clashes of theories settled by decisive experiments, these are, as Imre Lakatos put it, (historically corroborated) rational reconstructions: narratives we build much later, once a theory has eclipsed its predecessors and challengers. In practice, however, most science isn't really strictly deductive or inductive, but rather a messy process of fiddling with concepts and conjectures, crafting experiments and structured observations, interpreting the results, then going back to our conjectures.

A tempting thought, then, is that, in morals as in science, we need to reflect on how we arrive at our beliefs *and* how they hang together, rather than obsessing over deductive consistency and seeking unassailable foundations. If the real worlds of science look like reflective equilibrium-seeking, then doesn't that count in favour of coherentist justification in ethics and political philosophy?

The obvious counterpoint is that, at the end of the day, when I report results from a scientific experiment, I'm reporting on facts about the world that I've discovered. When I report on my deep moral convictions, I'm not doing the same thing at all. This distinction is buttressed for critics by the thought that, in science, there are certain privileged epistemic procedures (proper experimental design, valid causal statistical inference) that justify our faith in those observations. And (it is supposed that) there are decisive philosophical justifications for that epistemic privilege.

One way to rebut this claim, and recover the morals-science analogy, is to attack the second point just mooted: to argue that, in fact, those "observation reports" in science are much less secure than advocates think. These arguments attack the philosophical grounds for epistemic privilege in science, and that terrain is familiar, even in popular culture: science as culturally specific - imperialist, even - just one of many possible 'ways of knowing', and is given such deference out of the vagaries of political power and cultural chauvinism.

While that approach certainly has some plausibility (science does emerge out of a specific set of historical pathways, and has been bound up with all manner of arrogant and exclusionary beliefs) I don't think this approach ultimately succeeds, because it is too concessive about the possibility of moral justification, and because, at the end of the day science is an astonishingly useful 'way of knowing' the world, and for just the reasons that philosophers use to justify that epistemic privilege accorded to the results of well-designed experiments and statistical analyses.

Instead, I argue that the morals-science analogy holds, because at its most persuasive, Rawls's approach looks more like science. Rather than questioning the privileged status of observation reports in science, I argue that in fact there are similar reasons to privilege some kinds of moral intuitions in talking about legitimacy and justice.

Rawls and Overlapping Consensus: Modeling Full Reflective Equilibrium

Richard Lohse (KIT)

In his *Political Liberalism*, Rawls argues that a society can only be stably just if its citizens by and large agree on fundamental issues of justice. At the same time, he acknowledges the fact of reasonable pluralism in liberal democracies. To bring both together, Rawls introduces the concept of an overlapping consensus: The different comprehensive doctrines in a liberal society may overlap on a political conception of justice, e.g. justice as fairness. Optimally, this consensus can come about by using the method of reflective equilibrium: citizens carefully consider alternative conceptions and their arguments and, after due reflection, come to overlap on one conception of justice. In this case, they reach a reflective equilibrium that is both wide and general, or full. Rawls does not argue in detail for the possibility of a full reflective equilibrium. This paper picks up the thread. Using the formal model of reflective equilibrium recently developed by Beisbart, Betz and Brun, the paper distinguishes different kinds of conditions that bear on the likelihood of a full reflective equilibrium. It presents the design of a simulation study that focuses on the influence of certain structural features of the dialectical situation in a society. First results of the study are discussed.

Are Understanding and Reflective Equilibrium Collective Achievements?

Federica I. Malfatti (Innsbruck)

In ordinary speech, we regularly attribute understanding to collective entities – such as groups and communities. But what do we mean exactly when we claim that, say, the scientific community understands climate change? Do we mean that all or most members of this community understand the relevant phenomena? Or do we also mean something else – e.g., that the community, as a *community*, understands? Can a collective entity have an epistemic life on its own when it comes to understanding, or does the understanding of a collective entity always reduce to the understandings of its individual members? The overarching aim of this paper is to show that group or community understanding is a distinctively collective epistemic phenomenon that does not (always) result from the sum of individual understandings. I'll start by sketching a model of understanding phenomena inspired by the work of Elgin (2017) and Baumberger and Brun (2021). The notion of reflective equilibrium will play a central role in the model. I'll then formulate two divergence arguments to show that the epistemic life of a group or community and the life of its individual members can diverge, when it comes to understanding phenomena: a group can understand a phenomenon, while arguably none of its individual members understand it.

Towards a Bargaining Model of Reflective Equilibrium

Marina Moreno and Adriano Mannino (LMU Munich)

Reflective equilibrium (RE) is endorsed by many moral philosophers as a method for normative and applied ethical theorising. A core purpose of the method is understood to consist in recognising and resolving conflicts between initially credible commitments. Standard ways to model the process of conflict resolution make use of aggregative principles, i.e. principles directing the reasoner to weigh up the relative strengths of the commitments and the resulting systematising theories, and to then adopt the overall strongest commitments and theories as the new equilibrium state. We argue that aggregative principles are not necessarily adequate in this context. Rather, conflicts between moral commitments may sometimes – perhaps often – be adequately resolved via a process of rational bargaining. That is, the individual commitments are to be modelled as the functional analogues of individual agents with conflicting preferences, and the respective systematisation corresponds to the solution of some bargaining game between them. In support of this, we offer a theoretical argument and point to the disadvantages of exclusively aggregative approaches to RE. We illustrate our bargaining-based approach with trolley cases as well as cases to do with the just distribution of medical resources.

Is Reflective Equilibrium Really a Method?

Tanja Rechnitzer (Hannover)

While reflective equilibrium (RE) is often cited as a central method of philosophy, there are reasons for severe doubts about its status as a method. Firstly, in stark contrast to the number of claims to be using RE, explicit applications of the method are difficult to find in the philosophical literature.

Secondly, RE is the target of serious objections, e.g., that it is too vacuous to provide any useful guidelines or real constraints. This raises the question whether RE can really be applied in a conscious and helpful way – in short, is reflective equilibrium really a method?

I tackle this question in two steps: First, I clarify what it means to be a method by making the difference between method and methodology. This allows me to argue that parts of the objections rest on unrealistic expectations about methods. Second, I take a closer look at RE, discussing to what degree it can meet reasonable expectations for a method. Ultimately, I argue that talking about RE as a method often generates false expectations. We should distinguish explicitly between methodology and method and conceive of RE primarily as a methodology.

Charlie Harry Smith (Oxford)

Digitising Reflective Equilibrium

The method of reflective equilibrium is overdue a twenty-first century update. Despite its apparent popularity, few theorists seem to ever follow the method to completion, and even fewer thoroughly, openly and transparently publish their attempts to do so in print. This paper proposes the digitisation of reflective equilibrium as a solution. Inspired by the global open science movement, I advocate coupling a novel, digital implementation of the method with new publication norms that can capitalise on the reproducibility of digital data. I make three main claims: that digitising will make it easier to a) methodically construct, b) widely disseminate, and c) thoroughly critique equilibria. I also provide practical guidance where possible. Altogether, I argue that embracing digital reflective equilibria as standard will not only help theorists to better realise the method's latent theoretical potential in practice, but also greatly extend its value as a justificatory device in academic discourses.

Murray Smith (Kent)

Reflective Equilibrium, Aesthetic Judgement, and the Reflective Afterlife of Art

What bearing does reflective equilibrium have on our appreciation of artworks? In this paper I will explore this broad, framing question through two more specific questions. The first of these concerns the extent to which our immediate engagement with works of art is characterised by the process of reflective equilibrium; the second concerns the longer 'afterlife' of this immediate engagement, focussing on the manner and extent to which the knowledge that we gain from an artwork is integrated with the knowledge we derive from other sources. In relation to the first question, I argue that insofar as arriving at an aesthetic judgement of a work involves a to-and-fro between the particular details of the work and the artistic categories that it evokes, there are good reasons to think that a core aspect of engaging with an artwork is indeed characterised by (a species of) reflective equilibrium. In relation to the second question, I explore the hypothesis that the cognitive functions and epistemic goods of art are distinctive but indivisible from our knowledge and understanding as a whole. The reflective afterlife of artworks in this sense is characterised by reflective equilibrium.

The Role of Higher-Order Evidence in the Method of Reflective Equilibrium

Folke Tersman (Uppsala)

A common concern about the method of reflective equilibrium is based on the claim that people are likely to end up with radically divergent systems of moral beliefs due to differences that do not exclude that they all follow the method flawlessly. I examine that claim through considering what the method plausibly says about how inquirers should accommodate disagreement. Information to the effect that others have beliefs that conflict with one's own is sometimes referred to as "higher-order evidence", as it is supposed to pertain primarily to the justification rather than the correctness of the beliefs. The discussion in the paper speaks to the more general question of the role of such evidence in the method.

Equilibration: Art or Science?

Kenneth Walden (Dartmouth)

Bringing a system into reflective equilibrium is a process, and we might wonder: can this process be reduced to a set of comprehensive and precise rules? That is, can we hope to find a set of rules that will take a disequilibrated system as an input and yield a unique output? I argue against this possibility on the basis of two arguments. The first is by underdetermination: we should not expect to find the

materials to ground the appropriateness of set of rules sufficiently fine-grained and comprehensive to do this. The second is a version of the “Achilles and the Tortoise” argument: any rules that might perform this task ought themselves be incorporated into the system that is to be equilibrated, but that leaves the equilibration itself as a further task. If I am right about this, then how is equilibration accomplished? My suggestion is that it has more in common with our creative endeavors, especially the production of art, than we might have thought. The equilibration of a science succeeds not because it conforms to some antecedent standard of coherence but because it exemplifies an original and compelling conception of order.