Table of Contents What's New Random Entry Chronological Archives About Editorial Information About the SEP Editorial Board How to Cite the SEP Special Characters Advanced Tools Contact Support SEP Support the SEP PDFs for SEP Friends Make a Donation SEPIA for Libraries Entry Navigation **Entry Contents** Bibliography Academic Tools Friends PDF Preview Author and Citation Info Back to Top Reflective Equilibrium

Stanford Encyclopedia of Philosophy

Menu Browse

Many of us, perhaps all of us, have examined our moral judgments about a particular issue by looking for their coherence with our beliefs about similar cases and our beliefs about a broader range of moral and factual issues. In this everyday practice, we have sought "reflective equilibrium" among these various beliefs as a way of clarifying for ourselves just what we ought to do. In addition, we may also have been persuading ourselves that our conclusions were justifiable and ultimately acceptable to us by seeking coherence among them. Even though it is part of our everyday practice, is this approach to deliberating about what is right and finding justification for our views defensible?

First published Mon Apr 28, 2003; substantive revision Fri Oct 14, 2016

Viewed most generally, a "reflective equilibrium" is the end-point of a deliberative process in which we reflect on and revise our beliefs about an area of inquiry, moral or non-moral. The inquiry might be as specific as the moral question, "What is the right thing to do in this case?" or the logical question, "Is this the correct inference to make?" Alternatively, the inquiry might be much more general, asking which theory or account of justice or right action we should accept, or which principles of inductive reasoning we should use. We can also refer to the process or method itself as the "method of reflective equilibrium."

The method of reflective equilibrium can be carried out by individuals acting separately or together. In the latter case, the method is dialogical and agreement among participants may or may not be accompanied by a search for coherence. We shall be focused on the method when it seeks coherence among beliefs, for an avowal of agreement may well not include a real coherence of beliefs.

In what follows, we first give an overview of the method of reflective equilibrium and comment briefly on its history. We then discuss in more detail the evolution of the method and its role in the work of John Rawls. Against that background, we then remark on some of the controversy surrounding the claim that coherence among our moral or our logical beliefs in reflective equilibrium counts as a justification for them, including the challenge to including moral intuitions or facts about the world in such a justification. Finally, we discuss some implications the method has for work in ethics.

- 1. The Method of Reflective Equilibrium
- 2. Recent History
- 2.1 Origins in justification of logic
- 2.2 Development in Ethics and Political Philosophy
- 3. Distinguishing Narrow from Wide Reflective Equilibrium
- 3.1 Narrow Reflective Equilibrium
- 3.2 Wide Reflective Equilibrium
- 4. Criticisms of Reflective Equilibrium
- 4.1 Challenges to Moral Intuitions
- 4.2 Rejection of Rawls's Constructivism and Wide Reflective Equilibrium
- 4.3 Epistemological Criticisms of Reflective Equilibrium
- 5. Applications and Implications of Reflective Equilibrium

Bibliography

Academic Tools

Other Internet Resources

Related Entries

1. The Method of Reflective Equilibrium

The method of reflective equilibrium consists in working back and forth among our considered judgments (some say our "intuitions," though Rawls (1971), the namer of the method, avoided the term "intuitions" in this context) about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. The method succeeds and we achieve reflective equilibrium when we arrive at an acceptable coherence among these beliefs. An acceptable coherence requires that our beliefs not only be consistent with each other (a weak requirement), but that some of these beliefs provide support or provide a best explanation for others. Moreover, in the process we may not only modify prior beliefs but add new beliefs as well. There need be no assurance the reflective equilibrium is stable—we may modify it as new elements arise in our thinking (Schroeter 2004). In practical contexts, this deliberation may help us come to a conclusion about what we

ought to do when we had not at all been sure earlier. (Scanlon 2002). We arrive at an optimal equilibrium when the component judgments, principles, and theories are ones we are un-inclined to revise any further because together they have the highest degree of acceptability or credibility for us. An alternative account retains the importance of revisability and emphasizes the positive role of examining our moral intuitions, but rejects the appeal to coherentism in favor a treating our intuitive moral judgments as the right sort to count as foundational, even if they are still defeasible (McMahan 2000, Nichols 2012).

The method of reflective equilibrium has been advocated as a coherence account of justification (as contrasted with an account of truth) in several areas of inquiry, including inductive and deductive logic as well as both theoretical and applied philosophy. The key idea underlying this view of justification is that we "test" various parts of our system of beliefs against the other beliefs we hold, looking for ways in which some of these beliefs support others, seeking coherence among the widest set of beliefs, and revising and refining them at all levels when challenges to some arise from others. For example, a moral principle or moral judgment about a particular case (or, alternatively, a rule of inductive or deductive inference or a particular inference) would be justified if it cohered with the rest of our beliefs about right action (or correct inferences) on due reflection and after appropriate revisions throughout our system of beliefs. By extension of this account, a person who holds a principle or judgment in reflective equilibrium with other relevant beliefs can be said to be justified in believing that principle or judgment.

Because we are expected to revise our beliefs at all levels as we work back and forth among them and subject them to various criticisms, this coherence view contrasts sharply with a variety of many foundationalist approaches to justification. In ethics, some foundationalist approaches take some subset of our moral beliefs as fixed or unrevisable. Other foundationalists at least claim that some subset of our moral beliefs are immediately or directly justified (perhaps even "self-evident") or warranted (leaving aside the issue of revisability) and serve as the basis on which all other beliefs are justified. Still others take some subset of our beliefs as at least justifiable independently of any other moral beliefs, even if they are justifiable in light of either necessary or contingent views of the person or human nature or through appeals to the logic of moral discourse (Timmons 1987). Reflective equilibrium is unlikely to single out any such group of privileged or directly justified beliefs, distinguishing itself from all these forms of foundationalism. Rawls (1974) thought it highly unlikely, but not impossible, that moral principles could be formulated so compellingly (be "self-evident") that we would favor them and their consequences over all our previously held considered moral judgments; he thus leaves foundationalism a slim possibility, a point seized on by some who are willing to ignore Rawls's probability claim that such foundationalism is not probable at all. If we take this probability judgment seriously, then the warrant for a belief lies (with high probability) in its coherence with other beliefs and not in its resting on beliefs for which foundationalist claims are made.

Because it is not foundationalist in these ways, reflective equilibrium also avoids some other problematic distinctions or claims that are part of an effort to show how some beliefs can be directly justified or warranted. For example, some foundationalists view particular moral judgments as fixed; others might think it is our moral principles or some deeper theoretical beliefs from which such principles might be derived that are fixed and unrevisable. Some proponents of both approaches have even claimed that a moral sense or faculty reveals these directly justified beliefs to us. For others, we can discover the foundational beliefs in some deep moral belief structure that is revealed to us through a careful examination of moral judgments and that is arguably a priori. A considerable part of contemporary work in substantive ethics treats appeals to moral intuitions or considered judgments in this way (Thompson 1976, McMahan 2000, G.A. Cohen 2007). Some who work in this fashion cite Rawls (1974) when they embrace reflective equilibrium. We return to the topic of intuitionism again later in this entry.

In contrast, advocates of reflective equilibrium need tell no controversial stories about credentials for a special subset of beliefs that is directly justifiable. Where such advocates credit some initial beliefs with strong initial acceptance, as Rawls (Rawls 1971) does when he refers to some beliefs as initial "fixed points," the beliefs remain revisable; in any case, they are not taken to be points we believe independently of other moral and non-moral beliefs we hold, and so they have no special foundations (Harman 2003). This revisability of the initial beliefs and their dependence on other beliefs when reasons for them are requested means that no such special epistemological story has to be told about them.

As we shall see, however, an important point of controversy, especially in ethics, is not that reflective equilibrium allows for the revision of all moral judgments but, rather, that it involves giving some initial justificatory weight to them at all.

# 2. Recent History

## 2.1 Origins in justification of logic

This approach to the justification of rules of inductive logic—without the label "reflective equilibrium"—was proposed by Nelson Goodman in his classic Fact, Fiction, and Forecast (Goodman 1955). Goodman's idea was that we justify rules of inference in inductive or deductive logic by bringing them into reflective equilibrium with what we judge to be acceptable inferences in a broad range of particular cases. No rule of inference would be acceptable as a logical principle if it was not compatible with what we take to be acceptable instances of inferential reasoning. In this sense, our beliefs about acceptable rules of inference are constrained by the "evidence" provided by what we believe to be good or correct examples or instances of inferential reasoning. At the same time, we should correct or revise our views about particular inferences we initially might think are acceptable if we come to see them as incompatible with rules that we generally accept and refuse to reject because they, in turn, best account for a broad range of other acceptable inferences.

Some have criticized this account as giving too much weight to our actual inductive practices (Stich 1990, Kelly and McGrath 2010). Obviously, not all elements of the everyday reasoning practices of all individuals are justifiable. For example, many of us, to our chagrin, have had to confess committing the error of the gamblers fallacy in our own betting on games or on the events of life. Quite generally, psychological studies reveal widespread errors in reasoning in a broad range of contexts. More recently, others have suggested that reflective equilibrium is problematic as a form of justification of inductive reasoning because it is fragile as a method, allowing some features of our beliefs to trigger significant changes in the equilibrium they reach, and it provides inadequate assurance about its reliability, as a way of telling us what beliefs to replace with other beliefs (Harman and Kulkarni 2006).

Though Goodman thinks justification of our reasoning practices depends on what inferences we accept when we reason inductively and deductively, he is not simply seeking to systematize whatever inferences we happen to find people sometimes—unreflectively—make. Instead, he insists that practice can and should be corrected as we work back and forth from tentative principles to practice, revising where appropriate, presumably eliminating the sorts of inconsistencies that some psychological studies, and our everyday experiences, reveal. But, some critics ask (Siegel 1992), if practice should be revised, why should we view "fit" with practice as at all justificatory?

A more generous reading of Goodman's proposal would widen the reflective equilibrium he proposes to include some of the beliefs about standards for acceptable inference that logicians develop (though some think that wide reflective equilibrium does not overcome the reliability problem noted above). Such standards are themselves not independent of all inferential practice. They have been developed to reflect views about what counts as good practice in light of the kinds of inferences that people abandon when they are made aware of their inconsistency with other inferences they will not give up. It is against this wider set of beliefs, including the articulation of such standards, that we can identify some inferences as performance errors or otherwise deviant patterns and correct those practices. Some changes in belief in dialogical contexts may well be best explained by noting that some individuals grasp the insights of others. Individuals working with the method of reflective equilibrium may thus see the point of criticisms of previous views they had accepted. This is where some dialogical and individual uses of the method coincide. (We turn to the distinction between wide and narrow reflective equilibrium shortly.)

### 2.2 Development in Ethics and Political Philosophy

Despite the fact that the origins of reflective equilibrium (minus the name) lie in mid-twentieth century discussions about the justification of inductive logic, its principal development through the rest of the century lies primarily in ethics and political philosophy. Specifically, the method was given prominence (and the name by which it is known) by John Rawls's description and use of it in A Theory of Justice (Rawls 1971). (Rawls 1951 had much earlier articulated a slightly different version of the view.) Although accounts of the justification of empirical knowledge have been developed that share with reflective equilibrium its coherentist

approach, they generally do not make explicit use of the terminology of reflective equilibrium, and we shall not discuss them here. Instead, we concentrate on the use of reflective equilibrium in ethics and political philosophy, where it has been deployed and criticized.

Rawls (Rawls 1971) argues that the goal of a theory of justice is to establish the terms of fair cooperation that should govern free and equal moral agents. On this view, the appropriate perspective from which to choose among competing conceptions or principles of justice is a hypothetical social contract or choice situation in which contractors are constrained in their knowledge, motivations, and tasks in specific ways. Because this choice situation is fair to all participants, Rawls calls the conception of justice that emerges from this choice "justice as fairness." Under these constraints, he argues, rational contractors would choose principles guaranteeing equal basic liberties and equality of opportunity, and a principle that permitted inequalities only if they made the people who are worst off as well off as possible.

Instead of simply accepting whatever principles contractors would choose under these constraints on choice, however, Rawls imposed a further condition of adequacy on them. The chosen principles must also match our considered judgments about justice in reflective equilibrium. If they do not, then we are to revise the constraints on choice in the contract situation until we arrive at a contract that yields principles that are in reflective equilibrium with our considered judgments about justice. This restriction constitutes a further assurance that the outcomes of the deliberation about fair terms of cooperation in the choice situation (the Original Position) actually count as focusing on justice rather than on some other domain. In effect, the device of the contract must itself be in reflective equilibrium with the rest of our beliefs about justice. The contract helps us determine what principles we should choose from among competing views, but the justification for using it and designing it so that it serves that purpose must itself derive from the reflective equilibrium that it helps us achieve.

The method of reflective equilibrium thus plays a role in both the construction and justification of Rawls's theory of justice (Daniels 1996; Scanlon 2002). Its role in construction is an example of its use as a form of deliberation. Critics of Rawls's theory and his method of reflective equilibrium, especially utilitarians, challenge the prominence the method gives to moral judgments or intuitions. Building a theory—constructing it—out of such initial judgments is building it on easily discredited bases, for are not many of our beliefs just the result of historical accident and bias, even superstition? Indeed, some argue that there is no justificatory force to the contract itself if we "rig" the contract so that it yields principles that match our intuitions (Hare 1973). Despite these and other criticisms, defenders of the method have elaborated it and extended its use in broad areas of ethics. Later in this article we shall consider some of the criticisms of this method and some of its extensions in more detail. Before doing so, however, it is necessary to describe it more fully.

### 3. Distinguishing Narrow from Wide Reflective Equilibrium

# 3.1 Narrow Reflective Equilibrium

A reflective equilibrium may be narrow or wide (Rawls 1974). All of us are familiar with a process in moral deliberation in which we work back and forth between a judgment we are inclined to make about right action in a particular case and the reasons or principles we offer for that judgment. Often we consider variations on the particular case, "testing" the principle against them, and then refining and specifying it to accommodate our judgments about these variations. We might also revise what we say about certain cases if our initial views do not fit with the principles we grow inclined to accept.

Such a revision may constitute a moral surprise or discovery (Daniels 1996). Suppose, for example, that we are considering whether we should ignore age in the distribution of medical treatments. Many initially believe that age is a "morally irrelevant trait," just like race, and they would insist that rationing medical services by age is just as unacceptable as rationing by race. On considering a variety of cases, however, it might become apparent that we all age, although we do not change race. This difference means that the different treatment of people at different ages, if systematically applied over the lifespan, does not create inequalities between persons, as it would in the case of race. We might be led by this realization to think that age rationing might be acceptable under some conditions when race rationing would never be, and this would be a moral surprise for many who changed their view.

To the extent that we focus solely on particular cases and a group of principles that apply to them, and to

the extent that we are not subjecting the views we encounter to extensive criticism from alternative moral perspectives, we are seeking only narrow reflective equilibrium. Presumably, the principles we arrive at in narrow equilibrium best "account for" the cases examined. Others, however, may arrive at different narrow reflective equilibria, containing different principles and judgments about justice. Indeed, one such narrow equilibrium might be characterized as typically utilitarian, while another is, we may suppose, Kantian or perhaps Libertarian. As a result, we still face an important question about justification unanswered by the method of narrow reflective equilibrium: which set of beliefs about justice should we accept?

Because narrow reflective equilibrium does not answer this question, it may seem to be a descriptive method appropriate to moral anthropology, not a normative account of justification in ethics. In fact, Rawls (Rawls 1971) at one point suggested that arriving at the principles that match our moral judgments in reflective equilibrium might reveal our "moral grammar" in a way that is analogous to uncovering the grammar that underlies our syntactic ability as native speakers of a language to make judgments about grammatical form. In support of the analogy, some contemporary theorists who systematically examine our moral intuitions, often through hypothetical as opposed to real cases, believe they are uncovering an underlying moral structure of principles (see Kamm 1993), perhaps one that is a priori.

Uncovering a syntax, however, is a descriptive and not a justificatory task. Once we can identify the grammar or rules that best account for a person's syntactic competency, we do not ask the question, Should that person have this grammar? We are satisfied to have captured the grammar underlying a person's idiolect. In ethics and political philosophy, in contrast, we must answer that justificatory question, especially since there is often disagreement among people about what is right, disagreement that is not resolved simply by pursuing narrow reflective equilibrium.

In A Theory of Justice (1971, revised edition 1999), Rawls does not use the terminology of narrow and wide reflective equilibrium, an omission he remarks about with regret in Justice as Fairness: A Restatement (2001, p. 31). Still, he comments that seeking a reflective equilibrium that merely irons out minor incoherence in a person's system of beliefs is not really the use of the method that is of true philosophical interest in ethics, just as we saw it might not be in the case of justifying logical inferences. Rather, he says, to be of interest to moral philosophy, a reflective equilibrium should seek what results from challenging existing beliefs by arguments and implications that derive from the panoply of developed positions in moral and political philosophy (Rawls A Theory of Justice 2nd Edition, 1999, p. 43). Such a reflective equilibrium would be the response to considerable critical pressures on the original beliefs. This effort would have the character of searching deliberation about what is right. It is this much broader form of challenge that Rawls labels the method of wide reflective equilibrium.

### 3.2 Wide Reflective Equilibrium

# 3.2.1 A Theory of Justice

Rawls's proposal is that we can determine what principles of justice we ought to adopt, on full reflection, and be persuaded that our choices are justifiable to ourselves and others, only if we broaden the circle of beliefs that must cohere. Indeed, we should be willing to test our beliefs against developed moral theories of various types, obviously not all such views (as Arras 2007, Strong (2010) comment, Kelly and McGrath 2010), or we would never arrive at a conclusion, but at least against some leading alternatives. (We often do so in discussion or deliberation with others. How conclusive an argument or deliberation is may depend on what alternative views are considered or on who is included in the deliberation.) In effect, Rawls's (Rawls 1971) pair-wise comparison of justice as fairness with utilitarianism is just such an exercise at seeking wide reflective equilibrium. In a wide reflective equilibrium, for example, we broaden the field of relevant moral and non-moral beliefs (including general social theory) to include an account of the conditions under which it would be fair for reasonable people to choose among competing principles, as well as evidence that the resulting principles constitute a feasible or stable conception of justice, that is, that people could sustain their commitment to such principles. Rawls's argument is that justice as fairness, rather than utilitarianism, is what emerges in wide reflective equilibrium.

For example, the constraints on choice in Rawls's contract situation, such as the veil of ignorance that keeps us from knowing facts about ourselves and our specific preferences, require justification. Rawls needs to show

that these constraints are "fair" to all contractors (thus "justice as fairness" is his label for this procedural account of how we find out what is just.) To provide such justification, Rawls appeals to beliefs about the fundamental "moral powers" of "free and equal" agents (they can form and revise their conceptions of what is good and they have a sense of justice). He also appeals to the ideal of a well-ordered society in which principles of justice play a particular role as public principles for reconciling disputes. Rawls must even provide us with an account of "primary social goods," necessary if agents who do not know their own actual preferences are to decide what principles would be better for them to choose.

The device of the contract is thus in reflective equilibrium with certain background theories that themselves contain moral beliefs. These are theories, or crucial beliefs, about the nature of persons, the role of morality or justice in society, and beliefs about procedural justice. The contract is not simply based on uncontroversial assumptions about human rationality, although it describes a rational choice problem within the constraints it imposes; nor is it based on formal considerations about practical reasoning or the logic or semantics of moral discourse. If Rawls were trying to justify the structure of the contract by appeal to theories that themselves were completely non-moral, then he would be offering the kind of independent justification for the principles that would characterize them as foundational (Daniels 1996, Timmons 1987), so the claim that the background theories are themselves moral is part of the rationale for concluding that Rawls is clearly rejecting foundationalism. (Some philosophers, as noted earlier, argue that reflective equilibrium is compatible with foundationalism, citing Rawls in this regard though they ignore his probability judgment that such a compatibility exists in principle and is highly unlikely to be met.) Without acceptance of that wider circle of moral beliefs, Rawls's construction lacks support, and if that construction were modified, some of his arguments against utilitarianism would be weakened.

Our beliefs about justice are justified (and, by extension, we are justified in holding them) if they cohere in such a wide reflective equilibrium. Obviously, the method of wide reflective equilibrium is here only illustrated by appeal to the detail of Rawls's use of it. If we abstract from that detail, we see that there is a complex structure and interaction of beliefs, at many levels of generality, that bear on the construction of an account of justice. We shall later return to this point when we talk about some of the criticisms of method and about implications of this method for work in ethics.

## 3.2.2 Political Liberalism and 'justice as political'

In A Theory of Justice, Rawls seemed to think that all people might converge (but see Kelly and McGrath 2010) on a common or shared wide reflective equilibrium that included "justice as fairness," the conception of justice for which he argues. Wide reflective equilibrium thus played a role in the construction of the theory, helping "us"—any and all of us—to articulate its key features in ways that led to principles that matched "our" considered judgments. At the same time, wide reflective equilibrium constituted an account of justification. Shared agreement on that wide equilibrium would produce a well-ordered society governed by principles guaranteeing equal basic liberties, fair equality of opportunity, and the requirement that inequalities be arranged to make those who are worst off as well off as possible.

In his later work, Political Liberalism (Rawls 1993), Rawls abandons the suggestion that all people might converge in the same, shared wide reflective equilibrium that contains his conception of justice. This important change comes about because of what Rawls calls "the burdens of judgment." Complexity, uncertainty, and variation in experience lead human reason, when exercised under conditions of freedom, of the sort protected by the principles of justice as fairness, to an unavoidable pluralism of comprehensive moral and philosophical views. (The need to find agreement in the form of an overlapping consensus of views arises domestically under conditions of liberty and the burdens of judgment and is evidenced further by the different perspectives that we might get from a global view of issues.) This unavoidable fact of reasonable pluralism makes one key feature of justice as fairness untenable, namely the account Rawls gave of the stability of his preferred conception of justice. A further feature of Rawls's late work is his clear eschewal of any account of truth, so the fact that wide equilibrium is offered as an account of justification, not truth, makes it compatible with the refusal to discuss moral truth (see Little 1984).

Rawls had imposed three basic conditions on the principles of justice. First, they must be chosen over alternatives under conditions fair to all contractors. Second, what contractors choose must match "our" considered moral judgments and other beliefs in (wide) reflective equilibrium. But third, the principles must

comprise a feasible or stable conception of justice. The test for stability is to ask if people raised under this view would conform to it over time with less strain of commitment than other conceptions would face. In effect, passing the test shows it is worth adopting this view because it will not prove so fragile that it is not worth the effort to institutionalize.

The problem facing Rawls because of reasonable pluralism is that his earlier efforts to show justice as fairness was stable depended on special views about the importance of autonomy that would be held only by some people and could not be assumed to be shared. His argument for stability had turned on appealing to the good of autonomy in ways that might not be acceptable to people holding certain comprehensive views. Stability was not demonstrated.

To address this problem, Rawls recasts justice as fairness as a "freestanding" political conception of justice on which people with different comprehensive views may agree in an "overlapping consensus." The public justification of such a political conception involves no appeal to the philosophical or religious views that appear in the comprehensive doctrines that form this overlapping consensus. Instead, we might think of this process of working back and forth among the key shared ideas in the public, democratic culture and the articulated features of the political conception of justice as a political reflective equilibrium (Daniels 1996). (We can imagine the process being carried out by individuals, for example, as citizens thinking about an issue of basic justice or judges about a matter of constitutional essentials, or as a collective process, in public deliberation. As noted in Table 1 below, people with different comprehensive world views may draw the boundary between public and private slightly differently and not all questions about basic justice will be settled in a way that rules out ongoing disagreement. Thus disagreements about some aspects of reproductive rights or the separation of church and state may remain contested features within public reason.) The goal of this exercise of what Rawls calls "public reason" is the articulation of such a political reflective equilibrium.

Still, there is no convergence on a shared wide reflective equilibrium that contains the political conception of justice. The political reflective equilibrium is not a shared wide reflective equilibrium, for it avoids appealing to the broad range of beliefs that would be included in many wide reflective equilibria. Nevertheless, wide reflective equilibrium still plays a critical role in justification even when justice is political in this way.

For individuals to be fully justified in adopting the political conception of justice, the conception articulated in the political reflective equilibrium, they must incorporate it within a wide reflective equilibrium that includes their own comprehensive moral or religious doctrine. It will count as a reasonable view for them only if the public conception of justice coheres with their other (religious and philosophical) beliefs in wide reflective equilibrium. (Individuals within a religion may find a religious basis for accommodating their religion to concerns of public reason, though this outcome is unusual unless the religious hierarchy in their religion has found such accommodation as well.) Individuals, whatever groups and associations they belong to, must be able to embark on the deliberative process called wide reflective equilibrium that leads them to endorse, from their own perspective, the content of the political reflective equilibrium, including its injunctions about using only public reason in thinking about certain basic features of justice or the constitution.

We can imagine use of wide reflective equilibrium if we modify slightly Rawls image of an overlapping consensus in which a shared political conception is a point of convergence among otherwise different comprehensive world views. Convert each reasonable comprehensive world view—for example, a Kantian, Millian, or a Protestant religious view that endorses a notion of free faith—into a wide reflective equilibrium that is reached by all who hold that view. These distinct wide equilibria now share a module, the political reflective equilibrium, though each wide equilibrium will justify the features of that module in light of beliefs that may not be acceptable in other wide equilibria.

To help see what has happened to the content of the wide reflective equilibrium when we shift from the picture in A Theory of Justice to that in Political Liberalism, consider Table 1 below.

Table 1: The contents of wide reflective equilibrium in Theory and Liberalism (from Daniels 1996, p. 157).

Theory

Liberalism

Original Position; Principles of justice; free and equal agents; well-ordered society; procedural justice; general social theory [n.b. not freestanding module]

Original Position; Principles of justice; free and equal agents; well-ordered society; procedural justice; general social theory [n.b. freestanding module]

Philosophical arguments justifying key elements of justice as fairness, such as argument about autonomy needed to show congruence between good for individuals and what is just [n.b. share rationale]

Rationale for key elements of justice as fairness [n.b. rationale specific to each wide reflective equilibrium]

Account of reasonableness and burdens of judgment

Rationale for boundary between public and nonpublic; specifics of boundary.

Rest of moral and religious views

Rest of moral and religious views

Considered moral judgments, all levels

Considered moral judgments, all levels

The first row of the table indicates the content of justice as fairness, including the background views that support the construction of the original position. This "module" is contained in both versions of a wide reflective equilibrium that justifies justice as fairness. In Theory, as the second row indicates, shared philosophical arguments justify key elements of the module, but in Liberalism, the rationale for each of these elements must derive from the distinctive features of the different comprehensive views. Thus, as Cohen suggests (1994, p. 1527), a Kantian, a Millian, and a religious person who believed in free faith might all support, but for quite different reasons, the idea that agents were free in the sense of being capable of forming and revising their conceptions of the good life.

As the third and fourth rows suggest, full justification of a political conception of justice requires acceptance of an account of the burdens of judgment that explain why pluralism is a fact of life. A comprehensive view capable of reaching accommodation with that pluralism and with the political reflective equilibrium it shares with other reasonable views must also develop a rationale for the specific way in which it draws a boundary between public and nonpublic spheres. Different comprehensive views, however, might vary in just how they draw that public/private boundary and there may be persistent controversy about how permeable it is. Nevertheless, the freestanding public conception of justice as fairness, for example, will be fully justified for anyone in any of these different wide reflective equilibria.

It would be natural to object that if we require that full justification for individuals includes appeal to the varied religious and philosophical beliefs they hold, then overlapping consensus is indeed harder to achieve. Remember, Rawls has abandoned a shared wide reflective equilibrium for just such reasons. Overlapping consensus is possible, however, because groups sharing comprehensive views modify the content of their comprehensive views over time in order to cooperate within shared democratic institutions. This process, on Rawls's view, involves philosophical reflection, which often draws on the complex resources of a tradition of thought in which disagreements had flourished. It also crucially depends on the moderating influence of living under democratic institutions that are governed by the shared conception of justice. The effect of both institutions and reflection about internal disagreements about doctrine is that reasonable comprehensive world views can find room, from within their own perspectives, for a wide reflective equilibrium that includes the elements of public reason (the political reflective equilibrium) and a willingness to engage in public methods of justification for them.

In sum, even though pluralism requires that we refrain from appealing to comprehensive world views in certain areas of political deliberation with others, wide reflective equilibrium remains at the center of Rawls account

of individual moral deliberation about justice. It survives as the coherentist account of "full justification" he defends.

### 4. Criticisms of Reflective Equilibrium

Key criticisms of the method of reflective equilibrium have challenged the role it ascribes to moral intuitions, Rawls's incorporation of empirical facts (e.g., about human nature, such as the role of incentives in motivation), and its coherentism. We take these up in turn.

#### 4.1 Challenges to Moral Intuitions

Rawls does not talk about moral intuitions as the starting material for the method of reflective equilibrium; instead he talks about considered moral judgments. Some commentators think the whole issue of the relation between the method of reflective equilibrium and moral intuitions is resolved by noting that Rawls does not identify considered moral judgments with an appeal to moral intuitions. But this ignores the criticism that can be focused on considered moral judgments as well as on moral intuitions. Central to the method of reflective equilibrium in ethics and political philosophy is the claim that our considered moral judgments about particular cases carry weight, if only initial weight, in seeking justification. This claim is controversial. Some of the most vigorous criticism of it has come from utilitarians, and it is instructive to see why.

A traditional criticism of utilitarianism is that it leads us to moral judgments about what is right that conflict with our "ordinary" moral judgments. In response, some utilitarians accept the relevance of some of these judgments and argue that utilitarianism is compatible with them. Thus Mill argued for a utilitarian foundation for our beliefs about the importance of individual liberty. Some utilitarians have even argued that key features of our "common sense morality" approximates utilitarian requirements and we have acquired these beliefs just because they do, unconsciously, reflect what promotes utility; they reflect the wisdom of a heritage.

An alternative utilitarian response to the claim that utilitarianism conflicts with certain ordinary moral judgments is to dismiss these judgments as pre-theoretical views—whether they are referred to as "intuitions" or "considered judgments" that probably result from cultural indoctrination and thus reflect superstition, bias, and mere historical accident. On this view, neither moral intuitions nor judgments should have evidentiary credentials and should play no role in moral theory construction or justification. Indeed, the prominent 20th-Century utilitarians Richard Brandt (Brandt 1979) and Richard Hare (Hare 1973) argued against Rawls, simply making "coherent" a set of beliefs that have no "initial credibility" cannot produce justification, since coherent fictions are still only fictions. Indeed, the conditional that Rawls describes under the process that we solicit considered moral judgments, namely that people be calm and have adequate information about the cases, do not by themselves do anything to assuage the utilitarian worries. Brandt (Brandt 1990) reaffirms his early criticism when he claims that considered judgments lack "evidential force" regarding a moral order and therefore coherence in reflective equilibrium has only a kind of persuasiveness that comes from coherence among many elements being more convincing than the conviction that comes from any of its parts.

This criticism has some—but not decisive— force, since two standard ways of supplying credentials for initial judgments are not available. One traditional way to support the reliability of these judgments or intuitions is to claim, as 18th century theorists did, that they are the result of a special moral faculty that allows us to grasp particular moral facts or universal principles. Modern proponents of reflective equilibrium reject such mysterious faculties. Indeed, they claim moral judgments are revisable, not foundational.

A second way to support the initial credibility of considered judgments is to draw an analogy between them and observations in science or everyday life. For example, what counts as observational evidence in science depends on theory, and theory may give us reasons to reject some observations as not constituting counter-evidence to a scientific law or theory. In this way we might see an analogy between the revisability of moral judgments and observations.

Developing this analogy, however, seems to require that we also tell some story about why moral judgments are reliable "observations" about what is right. The foundationalist view of moral intuitions held by McMahan (2000) and perhaps other recent theorists who draw extensively on moral intuitions in their work in substantive ethics also demands some story about their reliability and so does not avoid this difficulty, even if the view

also allows for defeasibility of some of the foundational intuitions. Perhaps we might need something like the causal story that some theorists of knowledge offer to explain the reliability of observations. Since no such story is forthcoming, opponents argue, proponents of reflective equilibrium must reject the requirement or give up the analogy. Proponents of reflective equilibrium might reject the requirement by suggesting it is premature to ask for such a story in ethics or by claiming that we can provide no analogous causal story for credible judgments we make in other areas, including mathematics or logic. So the utilitarian objection is not conclusive.

It might seem that the burden of argument has shifted to advocates of reflective equilibrium to show why "initial credibility" should be ascribed to moral judgments or intuitions. Defenders of reflective equilibrium may nevertheless reject this burden, arguing that critics, especially utilitarian critics, actually face the same problem. For example, Richard Brandt argues that "facts and logic" alone, and not moral intuitions, should play a role in moral theory construction and justification. On Brandt's view, we should choose moral principles when they are based on desires that have been subjected to maximal criticism by facts and logic alone (he calls it "cognitive psychotherapy"). We should avoid any appeal to moral intuitions that might infect this critique.

The desires that Brandt appeals to, however, are themselves shaped and influenced by the very same social structures that utilitarians complain have biased and corrupted our moral judgments. Nothing in the process of critique by facts and logic alone can eliminate this source of bias. If this utilitarian claim is right, it undercuts the suggestion that we can step outside our beliefs to arrive at some more objective form of justification. Instead, we may be better off recognizing that our process of critique—in the method of wide reflective equilibrium—is explicit about exposing the sources of bias and historical accident to criticism and revision, rather than fooling ourselves into thinking that desires are some sort of morally-uninfluenced layer of "facts."

## 4.2 Rejection of Rawls's Constructivism and Wide Reflective Equilibrium

Rawls claims that his view of justice is constructivist, meaning that he appeals to some general claims about the nature of persons as well as some empirical facts about human behavior or institutions as part of the justification for the principles of justice (or the choice situation that leads us to pick them). G.A. Cohen (2008) criticizes constructivism, so understood, as not being capable of articulating the content of what justice itself requires and instead being suited only for selecting rules of regulation for society. The problem, he argues, is that constructivism combines considerations of justice with other considerations (both empirical and moral). As a result, it does not tell us what justice itself requires. Specifically, he argues that if considerations of justice plus other values and some empirical facts suggest that we should adopt certain rules to regulate our institutions, those rules cannot be principles of justice, since other considerations than justice contribute to our thinking that we should adopt them. A possible reply to Cohen is that his view about constructivism collapses into his controversial metaethical claim that principles of justice cannot rest on general facts about human behavior or anything else, if the other values that we consider are appropriately focused on determining what we think about justice. Then the other considerations that continue to play a role are appeals to empirical facts.

To understand why the objection to constructivism collapses into Cohen's metaethical view, if any values relevant to choices in the Original Position are relevant to determining what justice requires, consider the following. Cohen claims that, if we appeal to reflective equilibrium as a justification of what justice requires, we need to strip away from it any appeal to empirical facts and include only moral intuitions about justice itself (Cohen 2007: p. 243, n. 19). Specifically, he is saying that we might choose the right account of justice itself through appeal to a narrow reflective equilibrium that avoids any empirical considerations, and this is true even if the account of justice that applies to our world does involve some relevant facts. This restriction on reflective equilibrium implies that we would have to revise the account of the role of wide reflective equilibrium that we earlier attributed to Rawls, specifically that it, and not narrow reflective equilibrium, plays a justificatory role in supporting justice as fairness. For example, we would have to abandon the argument made earlier that inadequate critical considerations play a role in narrow reflective equilibrium as compared with wide reflective equilibrium.

Rawls insists that the principles that emerge from the Original Position must fit with our considered judgments

about justice in wide reflective equilibrium. This constraint on the outcome of the Original Position is aimed at showing that these principles are really principles of justice. On Cohen's view this constraint could not show what Rawls intends—the outcome for him of the choices made in the Original Position are rules of regulation, not principles of justice. But if Cohen is wrong (as the objection we are examining implies) about a broader set of values than justice being in play in the choices made in the Original Position, then his objection to constructivism collapses into his metaethical view, for it is an objection to appealing to the facts that are part of the wide reflective equilibrium to which Rawls does appeal. Similarly, Rawls's overall account of justification makes the relative stability of a conception of justice (given facts about human nature and behavior) part of the justification for his principles of justice as fairness, but Cohen concludes that this role is inappropriate.

If Cohen is right, then wide reflective equilibrium is not relevant to choosing between Rawls's account of justice and some other theory (such as luck egalitarianism), because we must restrict ourselves to a different, much narrower, version of reflective equilibrium to justify principles of justice. It would be odd to make such a substantive issue depend on a metaethical claim that Cohen himself says is not a substantive claim about which principles of justice are to be believed or accepted.

## 4.3 Epistemological Criticisms of Reflective Equilibrium

Other criticisms of reflective equilibrium in ethics have focused on points familiar from discussions in epistemology more generally. One important complaint concerns the vagueness of the concept of coherence. If we simply take logical consistency as the criterion for coherence, we have much too weak a constraint. What account can we give of a stronger notion?

More has to be said about how some parts of the system of beliefs "support or explain" others than is provided in the account above. Critics of Rawls's requirement that the contract situation be adjusted, if necessary, to yield principles that are in reflective equilibrium with our judgments about justice complained that this was a "rigged" contract and that it did no justificatory work beyond reflective equilibrium. If, however, the moral judgments that play a role in discussions of fair process, the well-ordered society, and the moral powers of agents are somewhat independent of the considered judgments about justice, then we get the kind of independent support for the principles that add justificatory force. Little work has actually been done, however, to flesh out a stronger account of coherence in ethics as opposed to epistemology more generally, where coherentism has been defended and elaborated.

A second line of objection derives from anti-coherentist accounts of justification rooted in the theory of knowledge. Some insist, for example, that coherence accounts of justification cannot be divorced from coherence accounts of truth. On the other hand, Rawls view fits with the claim by some contemporary theorists of knowledge that a coherence account of justification is distinguishable from a coherence account of truth and defensible when so separated.

This separability of justification from claims about truth is important to Rawls, even more so in his later work than in A Theory of Justice. In his earlier work, Rawls never claimed that what emerged from his contract and other justificatory conditions were "truths" of justice. (Arras's (2007) conclusions emphasize the point that the method of reflective equilibrium does not assure us of truth—but the point was made explicit by Rawls himself decades earlier.) Rather, the method of reflective equilibrium suggested ways in which convergence might be achieved among those who began with disagreements about justice (of course, as Arras (2007) and Kelly and McGrath (2010) emphasize, the method of reflective equilibrium does not guarantee convergence either). By drawing attention to the many features of a comprehensive theory on which argument and evidence could be brought to bear, the method of wide reflective equilibrium held out promise that more convergence might result than if people had only considered judgments and principles to agree and disagree about. Convergence did not imply that truth was reached. A proponent of moral truth might still hope, however, that convergence could be taken as evidence for truth, just as it commonly claimed in non-normative areas of inquiry, contrary to the phenomenon that believing may yield seeing.

Once Rawls took the "burdens of judgment" and reasonable pluralism seriously, that is once he "politicized" justice in his later work, it became much less plausible to talk about convergence on a shared wide equilibrium that might contain moral truth. Comprehensive philosophical views, including those about moral truth, were

barred from playing a role in seeking the political reflective equilibrium involved in overlapping consensus. Even if we converged in an "overlapping consensus" on a conception of justice, individuals and groups would be able to claim they are fully justified in accepting that conception of justice only if it cohered with the perspectives of their distinct wide reflective equilibria. It became more important for Rawls to suggest how claims about justice could be "objective" without presupposing moral truth. These developments in Rawls's account thus appear to move him farther away from those who would seek to give a realist account of moral truth. Thus when Arras (2007) concludes that reflective equilibrium cannot be a theory of truth, he does not disagree with Rawls, as noted earlier. More challenging is his conclusion that reflective equilibrium fails to provide a basis for intersubjective agreement if we take seriously the importance of giving reflective equilibrium a global scope, since starting points are so different.

Arras's point may be a version of a claim about human rationality. A starting point for this line of argument is the complaint that people who have different starting points as their initial set of beliefs, say with different degrees of credence given to their beliefs, may arrive at different reflective equilibria points—a specific version of which is Arras's point about the global scope of claims of intersubjectivity. This possibility undercuts the suggestion that critical pressures alone deriving can produce convergence. A more general form of this worry is that the model of reflection involved in reflective equilibrium, in which we are to produce coherence among all our beliefs, overemphasizes and idealizes human rationality. The suggestion is that we would be better off presupposing a more minimal form of rationality. A specific version of this complaint is that the method involves an information burden that cannot be met. Yet another version of this criticism is that we should allow for much less "rationalist" forms of modification of our views, recognizing "conversion" experiences like those involved in "paradigm shifts" may affect any account of coherence. Another line of argument emphasizes the information burden of the method—how can we be sure all appropriate theoretical views have been addressed. The suggestion (Arras 2007) is that wide reflective equilibrium is not action guiding in that it does not tell us which views to save and which to revise and so it does not lead to justification or truth.

A full defense of reflective equilibrium as a method would require a more developed response to many of these lines of criticism than exists in the literature.

## 5. Applications and Implications of Reflective Equilibrium

Despite these criticisms, some philosophers have argued for a broader understanding of the relevance of the method of reflective equilibrium to practical ethics. In thinking about the course of right action in a particular case, we often appeal to reasons and principles that are notoriously general and lack the kind of specificity that make them suitable to govern the case at hand without committing us to implications we cannot accept in other contexts. This requires that we refine or specify the reasons and principles if we are to provide appropriate justifications for what we do and appropriate guidance for related cases. Philosophers who have focused attention on the importance of specification have drawn on the method of reflective equilibrium for their insights into the problem.

In practical ethics, especially bioethics, there has been a vigorous debate about methodology. Some argue that we must root all claims about specific cases in specific ethical theories, and the hard work is showing how these theories apply to specific cases. Others argue that we may disagree about many aspects of general theory but still agree on principles, and the hard work of practical ethics consists in fitting sometimes conflicting principles to particular cases. Still others argue that we must begin our philosophical work with detailed understanding of the texture and specificity of a case, avoiding the temptation to intrude general principles or theories into the analysis.

A grasp of the method of wide reflective equilibrium suggests a way around this exclusionary nature of this debate. Wide reflective equilibrium shows us the complex structure of justification in ethics and political philosophy, revealing many connections among our component beliefs. At the same time, there are many different types of ethical analysis and normative inquiry.

This suggests a more eclectic view of the debate about method. Work in ethics requires all levels of inquiry proposed by the disputants, not always all at once or in each case, but at some time or other. Sometimes it is true that we cannot resolve disputes about how to weigh conflicts among principles unless we bring more theoretical considerations to bear (these considerations need not involve comprehensive ethical theories).

Sometimes we can agree on relevant principles and agree to disagree on other theoretical issues, still arriving at agreement about the rightness of particular policies or actions and their justification in light of relevant reasons and principles. Sometimes we must see what is distinctive about particular cases and revise or refine our reasons and principles before we can arrive at understanding of what to do. Finally, and of the greatest significance, the method of wide reflective equilibrium should make it clear that work in ethical theory cannot be divorced from work in practical ethics. We must test and revise theory in light of our considered judgments about moral practice. The condescending attitude of many who work in ethical theory toward work in practical ethics thus is incompatible with what wide reflective equilibrium establishes about the relationship between these areas of ethical inquiry.

The method of wide reflective equilibrium makes it plausible that all such "methods" should be seen as appropriate to some tasks in ethics and are but parts of a more encompassing method.

# Bibliography

Arras, J., 2007, 'The Way We Reason Now: Reflective Equilibrium in Bioethics', in The Oxford Handbook of Bioethics, B. Steinbock (ed.), New York: Oxford University Press, pp. 46-71.

Brandt, R., 1979, A Theory of the Good and the Right, Oxford: Oxford University Press.

—, 1990, 'The Science of Man and Wide Reflective Equilibrium', Ethics, 100: 259–278.

Cohen, G.A., 2008, Rescuing Justice and Equality, Cambridge: Harvard University Press.

Cohen, J., 1994, 'A More Democratic Liberalism', Michigan Law Review, 92(6): 1506-43.

Daniels, N., 1979, 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', Journal of Philosophy, 76(5): 256–82; reprinted in Daniels, N., 1996, Justice and Justification: Reflective Equilibrium in Theory and Practice, Cambridge: Cambridge University Press, pp. 21–46.

—, 1996, Justice and Justification: Reflective Equilibrium in Theory and Practice, New York: Cambridge University Press.

DePaul, M.R., 1993, Balance and Refinement: Beyond Coherence Methods of Moral Inquiry, New York: Routledge.

Goodman, N., 1955, Fact, Fiction, and Forecast, Cambridge, MA: Harvard University Press.

Hare, R.M., 1973, 'Rawls' Theory of Justice', Philosophical Quarterly, 23: 144-55; 241-51.

Harman, G., 2003, 'Three trends in moral and political philosophy', Journal of Value Inquiry, 37: 415–25.

Harman, G., and Kulkarni, S., 2006, 'The Problem of Induction', Philosophy and Phenomenological Research, 52(3): 559–75.

Kamm, F., 1993, Morality and Mortality Vol. 1., Oxford: Oxford University Press.

Kelly, T. and McGrath, S., 2010, 'Is Reflective Equilibrium' Philosophical Perspectives, 24(1): 325–359.

Little, D., 1984, 'Reflective Equilibrium and Justification,' Southern Journal of Philosophy, 22(3): 373–387.

McMahan, J., 2000, 'Moral Intuition', in Blackwell Guide to Ethical Theory, H. LaFollette (ed.), Oxford: Blackwell, chap. 5.

Nichols, P., 2012, 'Wide reflective equilibrium as a method of justification in bioethics,' Theoretical medicine and Bioethics, 33(5): 325-341.

Rawls, J., 1951, 'Outline of a Decision Procedure for Ethics', Philosophical Review, 60(2): 177–97, reprinted in Collected Papers, 1999, pp. 1–19.

—, 1971, A Theory of Justice, 2nd Edition 1999, Cambridge, MA: Harvard University Press.

—, 1974, 'The Independence of Moral Theory', Proceedings and Addresses of the American Philosophical Association, 47: 5–22, in Collected Papers, 1999, pp. 286–302.

- —, 1993, 1996 Political Liberalism, New York: Columbia University Press, Paperback Edition.
- —, 1999, Collected Papers, Sam Freeman, (ed.), Cambridge MA: Harvard University Press.
- —, 2001, Justice as Fairness: A Restatement, Cambridge MA: Harvard University Press.

Scanlon, T.M., 2002, 'Rawls on Justification', in The Cambridge Companion to Rawls, S. Freeman (ed.), Cambridge: Cambridge University Press, pp. 139–167.

Schroeter, F., 2004, 'Reflective Equilibrium and Anti-theory', Noûs, 38(1): 110-134.

Sencerz, S., 1986, 'Moral Intuitions and Justification in Ethics', Philosophical Studies, 50: 77–95.

Siegel, H., 1992, 'Justification by Balance', Philosophy and Phenomenological Research, 52(1): 27–46.

Stich, S., 1990, The Fragmentation of Reason, Cambridge, MA: MIT Press.

Strong, C., 2010, 'Theoretical and Practical Problems with Wide Reflective Equilibrium in Bioethics', Theoretical Medicine and Bioethics, 31(20): 123–140.

Thompson, J.J., 1976, 'Killing, Letting Die, and the Trolley Problem', The Monist, 59: 204–217.

Timmons, M., 1987, 'Foundationalism and the Structure of Ethical Justification', Ethics, 97(3): 595-609.

van der Burg, W., and van Willigenburg, T., (eds.), 1998, Reflective Equilibrium: Essays in Honour of Robert Heeger. Dordrecht, Kluwer Academic Publishers.

Academic Tools

How to cite this entry.

Preview the PDF version of this entry at the Friends of the SEP Society.

Look up this entry topic at the Internet Philosophy Ontology Project (InPhO).

Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

[Please contact the author with suggestions.]

Related Entries

justification, epistemic: coherentist theories of | moral epistemology | original position | pluralism | public reason | Rawls, John | social contract: contemporary approaches to

Copyright © 2016 by Norman Daniels <ndaniels@hsph.harvard.edu>

Open access to the SEP is made possible by a world-wide funding initiative. The Encyclopedia Now Needs Your Support Please Read How You Can Help Keep the Encyclopedia Free

Browse

Table of Contents

What's New

Random Entry

Chronological

Archives

About

**Editorial Information** 

About the SEP

Editorial Board

How to Cite the SEP

Special Characters

Advanced Tools

Contact

Support SEP

Support the SEP

PDFs for SEP Friends

Make a Donation

SEPIA for Libraries

Mirror Sites

View this site from another server:

USA (Main Site) CSLI, Stanford University

Info about mirror sites

The Stanford Encyclopedia of Philosophy is copyright © 2020 by The Metaphysics Research Lab, Center for the Study of Language and Information (CSLI), Stanford University

Library of Congress Catalog Data: ISSN 1095-5054