

John Rawls (1921–)

NORMAN DANIELS

John Bordley Rawls, who developed a contractarian defense of liberalism that dominated political philosophy during the last three decades of the twentieth century, was born in Baltimore, Maryland. In 1939, he left his home town to attend Princeton. He served in the Pacific (1943–5), and returned to Princeton to receive his Ph.D. (“A Study in the Grounds of Ethical Knowledge”) in 1950. He taught briefly at Princeton, Cornell, MIT, and then, for thirty years, at Harvard. He married Margaret Warfield Fox, a painter, in 1949. They raised two sons and two daughters and have lived for many years in Lexington, Massachusetts.

Rawls’s enormous influence in philosophy, law, economics, and political science is largely traceable to his major work, *A Theory of Justice* (1971). According to one survey, it is one of the five most cited philosophical books of the twentieth century. In contrast to the dominant emphasis in twentieth-century ethics on the analysis of moral language and on topics in metaethics, *Theory* argued rigorously for substantive moral principles and discussed their implications for the design of basic social institutions. (See, e.g. ANSCOMBE, AYER, FOOT, HARE, MOORE, STEVENSON; cf. CHOMSKY and POPPER.) This normative stand encouraged other work on justice, as well as on other areas of applied ethics, and it explains the relevance of Rawls’s work beyond philosophy. Rawls’s influence is also the result of his dedication to teaching. He has trained many of the leading philosophers in ethics and political philosophy over several generations.

Rawls’s first publication, in 1951, “Outline of a Decision Procedure for Ethics,” though deriving from his Ph.D. thesis, expresses a lifelong theme in his work. It proposes a procedure for selecting and justifying ethical beliefs and principles from among the diverse views people hold. As he struggled with this problem over the next two decades, he narrowed the scope of his proposed solution. In *A Theory of Justice*, Rawls focuses only on a procedure for selecting among competing principles of justice, not moral principles quite generally.

In *Theory*, Rawls uses a hypothetical social contract (the Original Position) to argue for principles of justice different from the utilitarianism that has long dominated Anglo-American philosophy. Deliberating behind a “veil of ignorance” that blinds them to distinguishing and potentially-biasing facts about themselves, rational contractors choose principles that protect certain basic liberties, including the effective exercise of political liberties, guarantee fair equality of opportunity, and permit inequalities (measured

by an index of primary social goods) only if the inequalities work to make those who are worst off as well off as possible (the Difference Principle). Together these principles regulate the basic structure of society and produce a form of egalitarianism that Rawls calls “democratic equality.” Because these principles are chosen in a situation that is fair to all contractors, Rawls labels his view “justice as fairness,” by which he means procedural fairness.

In addition to being the rational choice of contractors, the principles must also meet two other conditions. They must match or cohere with “our” considered judgments about justice in (wide) “reflective equilibrium.” They must also be feasible in the sense that people raised in a society governed by them would find the system to be more stable, with less strain of commitment, than alternatives.

Following the publication of *Theory*, there was extensive critical response in philosophy journals and books, as well as in related fields. Rawls engaged actively with this critical literature over the next two decades. In the first few years after *Theory*, he defended his focus on the basic structure of society, his coherentist account of justification, his use of the primary social goods, his argument for the Difference Principle, the sense in which his view was fair to people with different conceptions of what is good, and he clarified the “Kantian interpretation” of justice as fairness (see 1999: chs 11–15). In 1980 Rawls published “Kantian Constructivism in Moral Theory” (the Dewey Lectures), in which he carefully described the details of the contract so that it represented the Kantian idea of free and equal agents who are rational and reasonable. On this constructivist view, there is no claim that the moral principles are “true” or represent a prior moral order.

Beginning in the mid-1980s, Rawls became concerned that he had underestimated the importance of the divergence among comprehensive moral and religious views that would emerge under the very conditions of liberty promoted by his theory. Could people with such divergent views have a stable agreement on a conception of justice? To answer this question and to accommodate the “reasonable pluralism” he thought unavoidable, Rawls revised his account of stability and political justification in papers leading up to and including *Political Liberalism* (1993). In *Liberalism*, he replaced Kantian constructivism with “political constructivism,” and the same ideas about free and equal citizens are used to construct a political conception of justice, again with no explicit claims about moral truth.

Rawls’s last major work is thus motivated by a central question about justification that evolved from his thesis and first publication: How can reasonable people with divergent moral and religious views come to agree on and abide by fair terms of cooperation?

Justice as fairness

The social contract

Rawls revives the social contract as a way to specify fair terms of social cooperation in the form of a hypothetical, not an actual or historical, agreement. The appeal to a contract embodies three main ideas. First, it is a form of *procedural* justice. When we do not have a prior principled agreement on what counts as fair or just, for example, about

how to reconcile concerns about liberty and equality and efficiency, we must rely on a procedure that is fair to all parties. We then can count the outcome of that procedure as fair or just.

Second, a fair procedure must embody features that are reasonable in light of the nature of the problem it addresses. In this case, the problem is to find principles of justice that “free and equal” persons can all agree provide the basis for a “well-ordered society.” As citizens or persons, we are free in the sense that we can form and revise a rational plan of life that specifies our conception of the good. We are equal in that we all normally have an adequate sense of justice, a disposition to seek and abide by terms of fair cooperation. A well-ordered society is one in which citizens accept and know that others accept the same principles of justice, and those principles govern its basic institutions.

Third, if properly designed, the contract situation represents an Archimedean Point. It stands outside the biased or self-interested beliefs we happen to hold, as well as the entrenched inequalities that may motivate them. From this standpoint, the contractors can leverage new agreements on points of controversy by building on relatively fixed or uncontroversial points.

Though we might understand that an actual contract binds those who make it, or those who implicitly consent to it, why think Rawls’s hypothetical contract tells us anything about what we ought to do? The answer must be that we share enough substantive moral agreement about our nature as free and equal persons, the goal of arriving at a well-ordered society, and the appropriateness of the design of the contract situation that we accept it as a procedural solution to the problem of justice.

The original position

The hypothetical contract involves “reasonable” constraints on contractors who must make a “rational choice” of principles of justice, presented as pair-wise comparisons. The contractors are limited in both knowledge and motivations, and thus they must not be confused with the “fully informed” bargainers or rational choosers who populate standard rational choice problems. They operate behind a “thick” veil of ignorance that blinds them to information about their age, race, gender, class position, the society they will enter and its position in history. They are also blind to their “rational plan of life” of conception of the good, including their system of moral and religious values. This thick veil assures that their choice of principles will not be affected by the self-interest that might come from knowledge of any of these facts about themselves. Contractors do, however, have general social knowledge or they would not be able to evaluate the choice of principles for their effects on well-being. Their motivations are also constrained. They are “mutually disinterested,” meaning they are concerned about their own well-being and the well-being of those in a generation or two either side, but they are not generally benevolent, malevolent, or envious.

Having blinded contractors to their own detailed or “thick” conception of the good, Rawls must provide them with a basis for determining how one set of principles will make them better off than another. Otherwise there is no basis for a rational choice of principles. Rawls introduces a set of “primary social goods,” containing rights and liberties, powers and opportunity, income and wealth, and the social bases of self-respect.

A weighted index of these objectively measurable goods is the basis for measuring the effects of alternative principles on well-being and for measuring inequalities between (representative members of) social groups. Consequently, the rational choice problem facing contractors is to decide which of two alternative principles under consideration leads to the highest index score for them.

Together with the requirement that all contractors have veto power over choices, the result of the “reasonable” constraints is to establish a baseline of equality that eliminates the influence of entrenched social inequalities.

Principles of justice

Rawls argues that his contractors would choose three principles of justice in preference to utilitarianism. His First Principle assures citizens they will have a set of equal basic liberties, including freedom of thought, expression, and association, security of the person, and rights of political participation. Rational contractors know that they may have fundamental moral and religious commitments, even if behind the veil they do not know exactly what they are. Once they reach some modest threshold of material well-being, having the liberty to pursue those commitments is not something they would trade for increments in income and wealth. Recognizing others and being recognized by them as political equals is an important social basis of self-respect as well.

So important are the recognitional aspects of the effective exercise of political participation rights that Rawls argues for special institutional protections of them. These institutional protections, such as public funding of elections, are intended to make sure that political participation rights are not merely formal but actually effectively exercisable by all, regardless of other inequalities. For these reasons, contractors would assure themselves these basic liberties directly through the First Principle rather than depending for them on the outcome of an uncertain utilitarian calculation.

Rawls’s Second Principle actually consists of two other principles. The fair equality of opportunity principle not only prohibits legal and quasi-legal barriers to opportunity, as would the weaker “formal” equality of opportunity principle, but also requires that positive steps be taken to mitigate the effects of social and economic contingencies on the developments of the talents and skills. These steps minimally include measures such as the provision of public education, but further early childhood interventions and family supports, such as day care, might be necessary to support fair equality of opportunity for both children and women. Both the First Principle and the fair equality of opportunity principle require certain kinds of equality. They are given priority over the Difference Principle, which allows certain inequalities.

The Difference Principle requires that inequalities, for example in income or wealth, be allowed only if they work to make the worst-off groups (and then the next worst off, etc.) as well off as possible. The idea is that it would be irrational for contractors to insist on equal shares of a small social produce if incentives would create a larger social product that could be divided so that all, even the worst off, benefit by getting more (according to the index of primary goods) than they would without incentives and inequalities. So far, this argument establishes only that it is irrational to disallow inequalities that advantage all. The Difference Principle, however, is very demanding

in that it requires inequalities to make the worst off as well off as possible (it is much more than “trickle down”). Crucial to Rawls’s argument for the Difference Principle is his claim that the very high stakes (lifetime prospects of well-being) and the great uncertainty imposed by the veil of ignorance mean that a “maximin” (maximize the minimum) principle is required as a principle of rational choice. Contractors are not permitted to gamble that they have an equal chance of being in any social position, a gamble that would make the principle of average utility preferable to them. Rather, the maximin principle requires they maximally protect the worst off through the Difference Principle. To make the Difference Principle seem less odd, Rawls also argues intuitively that both it and the fair equality of opportunity principle work to mitigate the effects of morally arbitrary social contingencies.

Utilitarians, Rawls notes, believe that in pursuing the aggregate welfare, the advantages of some outweigh the losses of others, much as the expenditure of effort in acquiring skills at one stage of life will be offset by greater rewards at another. Rawls’s principles together better recognize the “separateness of persons.” They afford stronger protections to individuals so that the advantages of social cooperation work more directly to the benefit of all.

Basic structure

Rawls intends his principles of justice to regulate the basic structure of society, that is, those major social institutions, such as the political constitution and the principal economic and social arrangements, that have “profound” effects on people because they distribute basic rights and duties and determine the division of advantages from social cooperation. In his later work, he explicitly includes the family in the basic structure. The principles of justice do not directly apply to the relationships individuals have with each other or in private associations. Rawls suggests there is an important division of moral responsibility: society assures that citizens’ needs are met through the principles of justice, which regulate basic institutions. Individuals are responsible for pursuing their rational plans of life within the constraints imposed by justice.

Primary social goods

Rawls’s rejection of utilitarian measures of well-being, such as welfare or desire satisfaction, in favor of an index of primary social goods, was challenged on several grounds. The index seems incomplete, for it fails to tell us who is worse off, the rich but ill person or the poor but well one. More generally, individual variations, such as those caused by disease or disability, would mean that individuals with the same primary social goods would actually have quite different capabilities. Some conclude that Rawls’s focus on the “resources” included in the index of primary social goods means he is concerned about the wrong “space”; egalitarians should focus more directly on individuals’ “opportunity for welfare or advantage” or on the capabilities or positive freedom they have.

In *Theory*, Rawls had made the simplifying assumption that all individuals were fully functional over a normal lifespan. This assumption invited these objections: that the index of primary goods was insensitive to important individual differences, such as disease or disability. By viewing disease and disability as impairments of the range

of opportunities open to people, it is possible to extend Rawls's theory to include problems of health and disease, and Rawls endorses such an extension in his later writings. Quite surprisingly, Rawls's three principles then constitute a fair distribution of the major "social determinants" of health, according to current work in the social sciences. In his later writings, Rawls also replies to the objection that the primary social goods might not be valued in the same way by people holding quite different conceptions of the good. He reformulates them as crucial "all purpose means" for meeting the needs of citizens. This reformulation, together with the extension to health care, makes his view converge more with that of some of his critics, for it suggests that justice as fairness is aimed at guaranteeing that all citizens' "needs" – as citizens – are met and that therefore they all have the capabilities to function – as citizens – as free and equal.

Wide reflective equilibrium

The principles that contractors choose must match our considered judgments about what is just in "reflective equilibrium." To achieve "reflective equilibrium," we work back and forth between our considered judgments about particular instances or cases, the principles that govern them, and the theoretical considerations that bear on accepting these considered judgments or principles, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. For Rawls, this means we should revise the constraints on choice in the Original Position until we arrive at a contract that yields principles that are in reflective equilibrium with our considered judgments. Thus, in his early work, the method of reflective equilibrium plays a role in both the construction and justification of his theory of justice.

There seems to be little justificatory force to achieving coherence or reflective equilibrium solely among principles and judgments about particular cases. Unless we think we have special knowledge of either the principles or judgments, which Rawls does not, such a *narrow* reflective equilibrium captures only what we happen to think is just. It does not show us that we are justified in holding those particular beliefs. Rawls believes we have no better method of justification than seeking a *wide* reflective equilibrium. This method broadens the field of relevant moral and non-moral beliefs to include both an account of the conditions under which it would be fair for reasonable people to choose among competing principles, and evidence that the resulting principles constitute a feasible or stable conception of justice, that is, that people could sustain their commitment to such principles. Our beliefs about justice are justified (and, by extension, we are justified in holding them) if they cohere in such a wide reflective equilibrium.

Central to the method of reflective equilibrium is the claim that our considered moral judgments about particular cases carry weight, if only initial weight, in seeking justification. Vigorous criticism of this claim comes from utilitarians, who denounce such "intuitions" as the results of historical accident and bias. Since, however, utilitarians allow individuals' desires or preferences to count in calculating what is good and right, and these desires are also affected by historical accident and bias, Rawls's openness about exposing moral judgments to comprehensive criticism in reflective equilibrium may be less harmful than the utilitarian approach.

In *A Theory of Justice*, Rawls seemed to think that all people might converge on a common or *shared* wide reflective equilibrium that included “justice as fairness.” We would be led to that equilibrium by philosophical argument about the various moral beliefs that contribute to the social contract approach, the details of the Original Position, and the arguments made within it. In his later work, Rawls modifies this view (see “Justification revisited” below).

Stability and feasibility

Principles of justice must not only be chosen by contractors and match our judgments in reflective equilibrium, but they must be more stable than alternative views. People raised in a well-ordered society must find that conforming to them involves less strain of commitment than conformity with alternatives. For example, the worst off arguably would find the strain of commitment less under the Difference Principle, which makes them as well off as possible, than they would under a utilitarian principle that simply maximized aggregate or average utility, assuming benefits to others compensated them for their losses. Because the autonomy exercised enabled by the principles of justice would be viewed as a good by people, Rawls thought his view stable. A growing respect for pluralism led Rawls to revise this argument for stability.

Justice as political

Burdens of judgment and reasonable pluralism

Reasonable people, especially under conditions in which they enjoy basic liberties, will tend to develop divergent comprehensive philosophical and religious views through which they assess what is valuable in life. By “reasonable,” Rawls means people who are concerned to live with others on fair terms, assuming that the others are so willing. Reasonable people also understand that to be fair the terms of cooperation must be ones that other free and equal persons can accept. Reasonable people will recognize that disagreements arise among them because of the “burdens of judgment.” These burdens include the conflicting and complex evidence that bears on issues, the disagreements about how to weight considerations, the vagueness of some of our concepts, the effects of the totality of a person’s experience on how she weights considerations, the multiplicity of normative considerations that are relevant and from which a selection must be made in any specific case. We are driven, Rawls concludes, to accept reasonable pluralism about many matters of importance. This is a basic fact of political life, and even among reasonable people we will find disagreements that threaten the original suggestion that philosophical argument could produce convergence on the same wide reflective equilibrium.

Overlapping consensus

Rawls addresses the problem of producing stable agreement despite reasonable pluralism by recasting justice as fairness as a “free-standing” *political* conception of justice. The key ideas out of which justice as fairness (or other, alternative reasonable

political conceptions of justice) are constructed, for example, the idea that citizens are free and equal, are now taken to be shared elements of our political life, that is, of a public, democratic culture. These ideas are already held or accepted by most people who share that culture, whatever other views they diverge on. In effect, it is not philosophy alone – aided by universal reason – that has led people to converge on these ideas, but a shared set of institutions and history. The appeal to a shared democratic culture, however, is not a concession to the “communitarian” critics of Rawls, who had complained that a shared conception of the good must unite people and form the basis for justice; instead, it is a way for Rawls to seek agreement among those who disagree about such views of the good, among other things.

Rawls suggests that we think of the political conception of justice as fairness as a “module” with its own internal principles, reasons, and standards of evidence. For example, justice as fairness includes the two principles of justice ordered in a particular way. Together these ordered principles, illuminated by the shared background ideas and publicly defensible standards of evidence and reasoning, specify the content of “public reason” as it is used to deliberate about matters of justice. This module should be complete: it should give “reasonable” answers to a broad range of questions about “constitutional essentials and basic questions of justice.”

These answers are “reasonable,” however, in light of the kinds of reasons to which the political conception is restricted. In effect, the justification for these answers only goes so far. It appeals only to reasons contained in the public view. Rawls calls this “*pro tanto* justification.”

People with divergent comprehensive moral and religious views can overlap in their acceptance of a conception of justice, the most reasonable of which Rawls thinks is justice as fairness. He draws an analogy to the same theorem’s being provable within different axiomatic systems. Nothing that turns on the comprehensive views plays a role in the public justification of the module. No claims about moral truth and no specific moral or philosophical views that are the distinctive features of such comprehensive views play a role.

Overlapping consensus is not a compromise or *modus vivendi* among competing groups that hold different moral conceptions. Public justification of the view must be for the “right reasons” and turn on the acceptability of the module and the ideas it rests on to those who hold those views.

Public reason

An idea that becomes central in Rawls’s later work is that of “public reason.” In *Political Liberalism*, Rawls argues for a rather restrictive view of public reason, attempting to restrict the introduction of religious and other views into public debate, especially by public officials and even in the thinking of citizens as they vote. This view was widely criticized, and Rawls adopts a more relaxed, “wide” view of public reason in his last paper on the topic, “The Idea of Public Reason Revisited.” In the wide view, deliberations about justice, especially by public officials, are governed by a “proviso.” Reasonable comprehensive moral or religious doctrines may be introduced into public political discussion at any time, and there may be good reasons for doing so, *provided*

that proper political reasons are also offered that are sufficient to establish the same conclusions. This proviso applies to public political culture and not to debate in the background culture, which has no such restriction on it.

Justification revisited

To say that a claim about what is just is justified solely by public reason (or *pro tanto*) is not yet to say that it is a fully justified belief for a particular person. The criterion for full justification ultimately remains acceptability in wide reflective equilibrium, and *pro tanto* justification deliberately refrains from seeking such deeper justification. By not seeking or alluding to deeper justification, *pro tanto* justification does not alienate those who have different reasons for accepting the module.

We obtain the greatest stability we can for a political conception of justice, Rawls argues, answering his central question in *Liberalism*, when there is the right type of “overlapping consensus” on it, that is, when there is overlapping consensus for the right reasons. People with different comprehensive moral views must justify for themselves, by their own lights, that is, in their own wide reflective equilibria, the acceptability of the module. Their rationales will thus differ in ways that reflect their other philosophical, moral, and religious beliefs. Some may insist, for example, that there is “moral truth,” others deny it. Some might see the principles of justice as forms of divinely given natural law; others may see it as a human construction. Ultimately, people are justified in accepting justice as fairness if it is acceptable to them in the different wide reflective equilibria they can achieve.

If there is *general* acceptance in this way of the module within the different “reasonable” comprehensive views in a society, Rawls says that we have “general” reflective equilibrium. General reflective equilibrium is not a shared wide reflective equilibrium – except for the overlap on the module.

Current applications and controversies

At the beginning of the twenty-first century, Rawls’s work continues to stimulate extensive discussion in several very active fields of political philosophy, including the following:

The family and feminism

In his latest writings, Rawls emphasizes that the family should be included in the basic structure and thus be regulated by principles of justice. At the same time, concerns about equality must be reconciled with concerns about the liberty of families to pursue religious or moral views that involve gender role differentiation affecting mothers and children. Rawls imagines robust institutional protections of women, including day care and other family support systems; at the same time, he imagines the debate about gender roles to be carried on in the background culture and not through intrusions into the family. Nevertheless, Rawls’s emphasis on principles of justice should be contrasted with virtue-based feminist approaches to ethics.

Egalitarianism

A substantial body of egalitarian literature has arisen that challenges Rawls from various directions. One prominent view suggests that Rawls's intuitive arguments for "democratic equality," which appeal to the moral arbitrariness of social and natural contingencies, yield a more egalitarian view than is embodied in Rawls's principles. On this view, we are owed compensation for any deficit in opportunity for welfare or advantage that arises through no fault or choice of our own. Another egalitarian challenge is that the leeway Rawls allows to individuals to pursue incentives or to make selfish domestic choices will undermine the possibility of achieving optimal results, as judged by Rawls's own principles.

International justice

An early challenge to Rawls was that he failed to discuss what kinds of obligations of justice are owed across national boundaries; a related criticism is that his theory is too much wed to the idea of nation states. Rawls's late publications on "The Law of Peoples" extend his contractarian views to discuss human rights.

Democratic deliberation

The view of democracy that emerges in Rawls emphasizes political participation as a way of realizing our moral capabilities. Democracy is not simply a procedural method of achieving agreement that we must employ because we lack substantive agreement on various matters. Instead, Rawls provides a foundation for emphasizing the deliberative elements of democratic theory.

Bibliography

Works by Rawls

1971: *A Theory of Justice*, Cambridge, MA: Harvard University Press.

1993: *Political Liberalism*, New York: Columbia University Press. (Paperback edition (1996) contains a new introduction and "Reply to Habermas," from *Journal of Philosophy* 92/3 (March 1995).)

1999: *Collected Papers*, ed. S. Freeman, Cambridge, MA: Harvard University Press.

Works by other authors

Daniels, N. (ed.) (1975) *Reading Rawls*, New York: Basic Books.

Freeman, S. (ed.) (forthcoming) *Companion to Rawls*.

Reath, A., Herman, B., and Korsgaard, C. M. (eds.) (1997) *Reclaiming the History of Ethics: Essays for John Rawls*, New York: Cambridge University Press.

Symposium on Rawlsian Theory of Justice: Recent Developments, *Ethics* 99/4 (July 1989).