
A THEORY OF JUSTICE

Revised Edition

JOHN RAWLS

THE BELKNAP PRESS OF
HARVARD UNIVERSITY PRESS
CAMBRIDGE, MASSACHUSETTS

© Copyright 1971, 1999 by the President and Fellows of Harvard College

All rights reserved

Printed in the United States of America

This book is a revised edition of *A Theory of Justice*,
published in 1971 by Harvard University Press.

Library of Congress Cataloging-in-Publication Data

Rawls, John, 1921–

A theory of justice / John Rawls. — Rev. ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-674-00077-3 (cloth : alk. paper). —

ISBN 0-674-00078-1 (paper : alk. paper)

1. Justice. I. Title.

JC578.R38 1999

320'.01'1—dc21 99-29110

formulate a conception of justice which, however much it may call upon intuition, ethical or prudential, tends to make our considered judgments of justice converge. If such a conception does exist, then, from the standpoint of the original position, there would be strong reasons for accepting it, since it is rational to introduce further coherence into our common convictions of justice. Indeed, once we look at things from the standpoint of the initial situation, the priority problem is not that of how to cope with the complexity of already given moral facts which cannot be altered. Instead, it is the problem of formulating reasonable and generally acceptable proposals for bringing about the desired agreement in judgments. On a contract doctrine the moral facts are determined by the principles which would be chosen in the original position. These principles specify which considerations are relevant from the standpoint of social justice. Since it is up to the persons in the original position to choose these principles, it is for them to decide how simple or complex they want the moral facts to be. The original agreement settles how far they are prepared to compromise and to simplify in order to establish the priority rules necessary for a common conception of justice.

I have reviewed two obvious and simple ways of dealing constructively with the priority problem: namely, either by a single overall principle, or by a plurality of principles in lexical order. Other ways no doubt exist, but I shall not consider what they might be. The traditional moral theories are for the most part single-principled or intuitionistic, so that the working out of a serial ordering is novelty enough for a first step. While it seems clear that, in general, a lexical order cannot be strictly correct, it may be an illuminating approximation under certain special though significant conditions (§82). In this way it may indicate the larger structure of conceptions of justice and suggest the directions along which a closer fit can be found.

9. SOME REMARKS ABOUT MORAL THEORY

It seems desirable at this point, in order to prevent misunderstanding, to discuss briefly the nature of moral theory. I shall do this by explaining in more detail the concept of a considered judgment in reflective equilibrium and the reasons for introducing it.²⁴

24. In this section I follow the general point of view of "Outline of a Procedure for Ethics," *Philosophical Review*, vol. 60 (1951).

Let us assume that each person beyond a certain age and possessed of the requisite intellectual capacity develops a sense of justice under normal social circumstances. We acquire a skill in judging things to be just and unjust, and in supporting these judgments by reasons. Moreover, we ordinarily have some desire to act in accord with these pronouncements and expect a similar desire on the part of others. Clearly this moral capacity is extraordinarily complex. To see this it suffices to note the potentially infinite number and variety of judgments that we are prepared to make. The fact that we often do not know what to say, and sometimes find our minds unsettled, does not detract from the complexity of the capacity we have.

Now one may think of moral theory at first (and I stress the provisional nature of this view) as the attempt to describe our moral capacity; or, in the present case, one may regard a theory of justice as describing our sense of justice. By such a description is not meant simply a list of the judgments on institutions and actions that we are prepared to render, accompanied with supporting reasons when these are offered. Rather, what is required is a formulation of a set of principles which, when conjoined to our beliefs and knowledge of the circumstances, would lead us to make these judgments with their supporting reasons were we to apply these principles conscientiously and intelligently. A conception of justice characterizes our moral sensibility when the everyday judgments we do make are in accordance with its principles. These principles can serve as part of the premises of an argument which arrives at the matching judgments. We do not understand our sense of justice until we know in some systematic way covering a wide range of cases what these principles are.

A useful comparison here is with the problem of describing the sense of grammaticalness that we have for the sentences of our native language.²⁵ In this case the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discriminations as the native speaker. This undertaking is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge. A similar situation presumably holds in moral theory. There is no reason to assume that our sense of justice can be adequately characterized by familiar common sense precepts, or derived from the more obvious learning principles. A

25. See Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, Mass., The M.I.T. Press, 1965), pp. 3–9.

correct account of moral capacities will certainly involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life; it may eventually require fairly sophisticated mathematics as well. Thus the idea of the original position and of an agreement on principles there does not seem too complicated or unnecessary. Indeed, these notions are rather simple and can serve only as a beginning.

So far, though, I have not said anything about considered judgments. Now, as already suggested, they enter as those judgments in which our moral capacities are most likely to be displayed without distortion. Thus in deciding which of our judgments to take into account we may reasonably select some and exclude others. For example, we can discard those judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside. All these judgments are likely to be erroneous or to be influenced by an excessive attention to our own interests. Considered judgments are simply those rendered under conditions favorable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain. The person making the judgment is presumed, then, to have the ability, the opportunity, and the desire to reach a correct decision (or at least, not the desire not to). Moreover, the criteria that identify these judgments are not arbitrary. They are, in fact, similar to those that single out considered judgments of any kind. And once we regard the sense of justice as a mental capacity, as involving the exercise of thought, the relevant judgments are those given under conditions favorable for deliberation and judgment in general.

I now turn to the notion of reflective equilibrium. The need for this idea arises as follows. According to the provisional aim of moral philosophy, one might say that justice as fairness is the hypothesis that the principles which would be chosen in the original position are identical with those that match our considered judgments and so these principles describe our sense of justice. But this interpretation is clearly oversimplified. In describing our sense of justice an allowance must be made for the likelihood that considered judgments are no doubt subject to certain irregularities and distortions despite the fact that they are rendered under favorable circumstances. When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit

his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral theory, the best account of a person's sense of justice is not the one which fits his judgments prior to his examining any conception of justice, but rather the one which matches his judgments in reflective equilibrium. As we have seen, this state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions (and the corresponding conception).

There are, however, several interpretations of reflective equilibrium. For the notion varies depending upon whether one is to be presented with only those descriptions which more or less match one's existing judgments except for minor discrepancies, or whether one is to be presented with all possible descriptions to which one might plausibly conform one's judgments together with all relevant philosophical arguments for them. In the first case we would be describing a person's sense of justice more or less as it is although allowing for the smoothing out of certain irregularities; in the second case a person's sense of justice may or may not undergo a radical shift. Clearly it is the second kind of reflective equilibrium that one is concerned with in moral philosophy. To be sure, it is doubtful whether one can ever reach this state. For even if the idea of all possible descriptions and of all philosophically relevant arguments is well-defined (which is questionable), we cannot examine each of them. The most we can do is to study the conceptions of justice known to us through the tradition of moral philosophy and any further ones that occur to us, and then to consider these. This is pretty much what I shall do, since in presenting justice as fairness I shall compare its principles and arguments with a few other familiar views. In light of these remarks, justice as fairness can be understood as saying that the two principles previously mentioned would be chosen in the original position in preference to other traditional conceptions of justice, for example, those of utility and perfection; and that these principles give a better match with our considered judgments on reflection than these recognized alternatives. Thus justice as fairness moves us closer to the philosophical ideal; it does not, of course, achieve it.

This explanation of reflective equilibrium suggests straightway a number of further questions. For example, does a reflective equilibrium (in the

sense of the philosophical ideal) exist? If so, is it unique? Even if it is unique, can it be reached? Perhaps the judgments from which we begin, or the course of reflection itself (or both), affect the resting point, if any, that we eventually achieve. It would be useless, however, to speculate about these matters here. They are far beyond our reach. I shall not even ask whether the principles that characterize one person's considered judgments are the same as those that characterize another's. I shall take for granted that these principles are either approximately the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments divide along a few main lines represented by the family of traditional doctrines that I shall discuss. (Indeed, one person may find himself torn between opposing conceptions at the same time.) If men's conceptions of justice finally turn out to differ, the ways in which they do so is a matter of first importance. Of course we cannot know how these conceptions vary, or even whether they do, until we have a better account of their structure. And this we now lack, even in the case of one man, or homogeneous group of men. If we can characterize one (educated) person's sense of justice, we might have a good beginning toward a theory of justice. We may suppose that everyone has in himself the whole form of a moral conception. So for the purposes of this book, the views of the reader and the author are the only ones that count. The opinions of others are used only to clear our own heads.

I wish to stress that in its initial stages at least a theory of justice is precisely that, namely, a theory. It is a theory of the moral sentiments (to recall an eighteenth century title) setting out the principles governing our moral powers, or, more specifically, our sense of justice. There is a definite if limited class of facts against which conjectured principles can be checked, namely, our considered judgments in reflective equilibrium. A theory of justice is subject to the same rules of method as other theories. Definitions and analyses of meaning do not have a special place: definition is but one device used in setting up the general structure of theory. Once the whole framework is worked out, definitions have no distinct status and stand or fall with the theory itself. In any case, it is obviously impossible to develop a substantive theory of justice founded solely on truths of logic and definition. The analysis of moral concepts and the a priori, however traditionally understood, is too slender a basis. Moral theory must be free to use contingent assumptions and general facts as it pleases. There is no other way to give an account of our considered judgments in reflective equilibrium. This is the conception of the subject

adopted by most classical British writers through Sidgwick. I see no reason to depart from it.²⁶

Moreover, if we can find an accurate account of our moral conceptions, then questions of meaning and justification may prove much easier to answer. Indeed some of them may no longer be real questions at all. Note, for example, the extraordinary deepening of our understanding of the meaning and justification of statements in logic and mathematics made possible by developments since Frege and Cantor. A knowledge of the fundamental structures of logic and set theory and their relation to mathematics has transformed the philosophy of these subjects in a way that conceptual analysis and linguistic investigations never could. One has only to observe the effect of the division of theories into those which are decidable and complete, undecidable yet complete, and neither complete nor decidable. The problem of meaning and truth in logic and mathematics is profoundly altered by the discovery of logical systems illustrating these concepts. Once the substantive content of moral conceptions is better understood, a similar transformation may occur. It is possible that convincing answers to questions of the meaning and justification of moral judgments can be found in no other way.

I wish, then, to stress the central place of the study of our substantive moral conceptions. But the corollary to recognizing their complexity is accepting the fact that our present theories are primitive and have grave defects. We need to be tolerant of simplifications if they reveal and approximate the general outlines of our judgments. Objections by way of counterexamples are to be made with care, since these may tell us only what we know already, namely that our theory is wrong somewhere. The important thing is to find out how often and how far it is wrong. All theories are presumably mistaken in places. The real question at any given time is which of the views already proposed is the best approximation overall. To ascertain this some grasp of the structure of rival theories

26. I believe that this view goes back in its essentials to Aristotle's procedure in the *Nicomachean Ethics*. See W. F. R. Hardie, *Aristotle's Ethical Theory*, ch. III, esp. pp. 37–45. And Sidgwick thought of the history of moral philosophy as a series of attempts to state "in full breadth and clearness those primary intuitions of Reason, by the scientific application of which the common moral thought of mankind may be at once systematized and corrected." *The Methods of Ethics*, pp. 373f. He takes for granted that philosophical reflection will lead to revisions in our considered judgments, and although there are elements of epistemological intuitionism in his doctrine, these are not given much weight when unsupported by systematic considerations. For an account of Sidgwick's methodology, see J. B. Schneewind, "First Principles and Common Sense Morality in Sidgwick's Ethics," *Archiv für Geschichte der Philosophie*, Bd. 45 (1963).