

Swarthmore College

## Works

---

Philosophy Faculty Works

Philosophy

---

11-1-1972

### Review Of "A Theory Of Justice" By J. Rawls

Hans Oberdiek

*Swarthmore College*, [hoberdi1@swarthmore.edu](mailto:hoberdi1@swarthmore.edu)

Follow this and additional works at: <https://works.swarthmore.edu/fac-philosophy>



Part of the [Philosophy Commons](#)

Let us know how access to these works benefits you

---

#### Recommended Citation

Hans Oberdiek. (1972). "Review Of "A Theory Of Justice" By J. Rawls". *New York University Law Review*. Volume 47, Issue 5. 1012-1022.

<https://works.swarthmore.edu/fac-philosophy/118>

This work is brought to you for free and open access by . It has been accepted for inclusion in Philosophy Faculty Works by an authorized administrator of Works. For more information, please contact [myworks@swarthmore.edu](mailto:myworks@swarthmore.edu).

## BOOK REVIEWS

A THEORY OF JUSTICE. By John Rawls. Cambridge, Massachusetts: Harvard University Press. 1971. Pp. xv, 607. \$15.00 cloth, \$3.95 paper.\*

### I

#### INTRODUCTION

The contribution of twentieth-century moral philosophy to the solution of pressing social problems has been minimal. G.E. Moore's "naturalistic fallacy" and the application of A.J. Ayer's verifiability criterion of meaning convinced a generation of philosophers that ethics, understood as the formulation and defense of substantive moral principles, is impossible. Philosophers' only legitimate task, so it was thought, lay in elucidating the language of morality and the vocabulary of politics. Thus, in a well-known book intended to bring the reader to grips with the fundamental problems of ethics, the author declared, "Ethics, as I conceive it, is the logical study of the language of morals."<sup>1</sup> Ethics, paradoxically, seemed to be irrelevant to moral concerns. One need not deny that linguistic analysis has clarified and deepened our understanding of moral concepts and judgments to suggest that philosophers must do more. In *A Theory of Justice*, John Rawls demonstrates that one can address living moral, political and social issues sensitively and forcefully without sacrificing philosophic standards of clarity and rigor.

*A Theory of Justice* includes much more than its title suggests. In addition to elaborating and defending principles of social justice, moral sentiments, principles of moral psychology and the unity of self, the nature of goodness and happiness are discussed. While these and other subjects are not simply addenda to Rawls's theory, I shall concentrate solely on the theory of social justice developed in the first two-thirds of

---

\* Citations to *A Theory of Justice* are to the paperback edition.

<sup>1</sup> R. Hare, *The Language of Morals*, at v (1952). Hare is not alone in his preoccupation with *meta-ethical* questions—with questions *about* moral discourse, especially questions concerning meaning, the nature of moral judgments and methods of justifying or supporting moral judgments. Throughout the nineteen forties and fifties, substantive moral and political philosophy was virtually dead. Studies sharing Hare's conception of ethics are: K. Baier, *The Moral Point of View* (1958); P. Nowell-Smith, *Ethics* (1954); C. Stevenson, *Ethics and Language* (1943); and S. Toulmin, *Reason in Ethics* (1950). For an instructive essay, see W. Frankena, *Moral Philosophy at Mid-Century*, 60 *Philosophical Rev.* 44 (1951). Two questions have troubled writers on meta-ethics in recent years: (1) Does acceptance of a particular meta-ethical theory logically commit one to the acceptance of any particular normative theory, and the converse; and (2) is the distinction between normative and meta-ethical judgments sound? The first question concerns *independence*; the second concerns *neutrality*. One of the first books to break away from an exclusive concern with meta-ethics is M. Singer, *Generalization in Ethics* (1961). Curiously, Rawls, who is generous in citing his predecessors, never mentions Singer's book.

the book. Particular attention will be paid to Rawls's general approach to theory-construction in ethics, his contractualism and the novel analytic construct from which he derives his twin principles of justice. Only after examining Rawls's methodology will I consider the principles of justice themselves. So that we will not lose our bearings, however, I begin with a thumbnail sketch of Rawls's general argument.

All of us have intuitive beliefs regarding justice, though few of us have anything like a theory of justice. Because our intuitions conflict or otherwise fail us, we seek a theory which will systematize and deepen our conception of justice. In developing any theory, we cannot begin from scratch; we must begin with our considered judgments, judgments that any plausible principle must confirm. However, as our theory develops we may modify or even abandon some of our initial beliefs, especially those which are, upon reflection, confused, inconsistent or the consequence of prejudice. Yet, however far our theory takes us from these initial beliefs, we cannot begin without them. Our aim is to reach a state of "reflective equilibrium," a state in which our principles and judgments coincide and in which we know the principles to which our judgments conform and the premises of their derivation (cf. p. 20).

To guide our deliberations we may "imagine that those who engage in social cooperation choose together, in one joint act, the principles which are to assign basic rights and duties and to determine the division of social benefits" (p. 11). The "most philosophically favored interpretation" of this initial choice situation, called *the original position*, will show that two principles of justice would be agreed upon, the first taking precedence over the second:

First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others.

Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all (p. 60).

These principles apply primarily to the basic structure of society, *i.e.*, to the political constitution and to the principal economic and social institutions. They may or may not have wider application. Further, both principles assume full compliance; social problems arising out of partial compliance create special difficulties. However, only an ideal theory can provide a systematic basis for the solution of contemporary social problems; unless we understand the nature and aims of a perfectly just society, we will lack a satisfactory standpoint from which to evaluate existing institutions and their practices.

## II

### METHODOLOGY

#### A. *The Pragmatic Approach*

Rawls adopts a currently fashionable account of theory-construction in the sciences. Rather than theorize on a bedrock of indubitable truths,

as Cartesians would have us do, we are urged to begin with what we do not *in fact* doubt, *e.g.*, that water boils at 100° C. at sea level. Hypotheses are then formed to ground such beliefs and to shed light on a variety of theoretical and practical problems we wish to solve (such as why altitude affects the boiling point of water). At an early stage of theorizing, hypotheses usually give way to firmly held beliefs if conflicts between them occur; but as hypotheses become imbedded in our conceptual scheme because of their explanatory power and connections with other hypotheses, we often sacrifice cherished beliefs in order to save our theories. In short, theorizing is a dialectical, self-correcting process of mutual adjustment among observations, hypotheses and theories.

Physical theories are now deeply imbedded in our thinking, though this was not always so. Psychological and sociological theories, on the other hand, have significantly less standing, so we are reluctant to modify our intuitive beliefs about human nature and society to satisfy someone's pet hypothesis. Ethical theories have even less claim to our allegiance. Most of us, for instance, believe that racism and slavery are unjust; if any theory implies otherwise, so much the worse for it! What we seek is a theory that will *explain* the injustice of racism and slavery and at the same time provide guidance where our moral beliefs are vague, tentative or inconsistent. While it would be rash to suppose that any ethical theory could possess the power of physical or biological theories, Rawls believes that his own conception of justice as fairness is a distinct improvement over previous conceptions, including utilitarian conceptions.

One virtue of Rawls's pragmatic approach is that it narrows the gap between scientific and ethical theories. Briefly, a scientific theory must answer adequately serious objections to it, have good reasons in its favor and be superior to its rivals in terms of simplicity, fruitfulness, thoroughness and compatibility with other accepted theories.<sup>2</sup> These same criteria apply to ethical theories, albeit in a somewhat different fashion. Unless an ethical theory can answer serious objections, for example, it must be rejected. No doubt the nature of "serious objections" and "adequate answers" changes as we pass from science to ethics, but this should not blind us to the underlying unity of the critical processes involved or to the fact that the nature of "serious objections" and "adequate answers" also varies within the sciences themselves. Another virtue of Rawls's pragmatic approach is that defi-

---

<sup>2</sup> There is a cursory discussion of the criteria of acceptability in science and their applicability to ethics in R. Garner & B. Rosen, *Moral Philosophy*, ch. 1 (1967). For a somewhat lengthier discussion of criteria of confirmation and acceptability, see R. Harré, *An Introduction to the Logic of the Sciences* (1960), or C. Hempel, *Philosophy of Natural Science* (1966). Rawls himself traces his approach to the work of W. Quine, *Word and Object* (1960), and of M. White, *Toward Reunion in Philosophy* (1956). Quine and White, colleagues of Rawls at Harvard, trace their roots back to the pragmatism of Charles Sanders Peirce and John Dewey.

nitions are kept in their place.<sup>3</sup> Key concepts are introduced and explained only as needed and only within the context of the entire theory; they have little or no life away from their theoretical home. Definitions occur, if at all, only after one understands the role of the defined concept within the theory. Thus, theories and definitions stand or fall together, an insight not always acknowledged by linguistic analysts.

One may feel that the pragmatic approach is bound to generate circular theories, since one uses intuitive convictions to formulate principles that are then confirmed by showing that they account for these same convictions. Were this all that was accomplished, one's theory would indeed be circular; fortunately, Rawls strives for much more. He argues, for example, that adoption of his two principles reconciles freedom and reason, helps explain the place of moral sentiments (such as shame, guilt, resentment and indignation) in our lives, fixes the proper relationship between majority rule and minority rights and explains the grounds for civil disobedience and political obligation. Principles of justice established through Rawls's general pragmatic approach therefore take us beyond our initial beliefs both by providing guidance in matters where we lack firm convictions and by showing how apparently unrelated beliefs form a coherent system.<sup>4</sup>

### B. *The Social Contract and the Original Position*

Social contract theories are in general disrepute for reasons which, if not always sound, are at least well-known. Rawls's decision to cast his theory in contractarian language may therefore come as a surprise. However, he sidesteps many familiar objections by denying that the contracting parties he has in mind are negotiating to enter a particular society or to form a particular government. Rather,

---

<sup>3</sup> See G. Field, *The Place of Definition in Ethics*, in 3 *University of Bristol Studies, Studies in Philosophy* (1935), reprinted in *Readings in Ethical Theory* 92-102 (1st ed. W. Sellars & J. Hospers eds. 1952).

<sup>4</sup> The pragmatic approach espoused by Rawls is not without pitfalls and limitations. William James, for instance, thought that a coherent, fruitful and superior theory was true. It follows from James's analysis that the Ptolemaic theory of planetary motion was once true but is now false. Rawls avoids this mistake, though he does so in part because he ignores questions of truth altogether. This may or may not be a virtue of the contract approach. Even granting the usefulness of the pragmatic approach in science, one wonders whether *moral* judgments, beliefs and convictions can play the same role that reports of empirical observations play. This question arises not simply because agreement on moral matters is difficult to obtain or because of the complexity of human conduct, but because one needs to determine and justify the *status* of moral judgments. Skeptics will argue that Rawls offers no more than a theory of a-rational preferences. To put the question in Kantian terms, Rawls must show that moral judgments are *possible* before claiming that he has established an objective theory of social justice. Following Kant further, we may ask whether ethical theories must in fact justify initial moral judgments at all—perhaps we should begin instead with certain moral concepts (such as duty) that even conflicting initial beliefs embody.

the guiding idea is that the principles of justice for the basic structure of society are the object of the original agreement. They are the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association. These principles are to regulate all further agreements; they specify the kinds of social cooperation that can be entered into and the forms of government that can be established. This way of regarding the principles of justice I shall call justice as fairness (p. 11).

Much of *A Theory of Justice* is devoted to explaining and defending the content of this passage.

Classical contract theorists had employed the notion of a state of nature to show what life would be like were men to live outside of society and to argue from that hypothetical state of affairs that certain features must characterize a just and well-governed society. Because Hobbes, Locke and Rousseau offered such radically divergent descriptions of life in a state of nature, it is not surprising that they defended equally divergent political principles and institutions. Although the concept of an *original position* resembles a state of nature in some respects, it differs in that it does not even pretend to describe a possible historical situation—the conditions defining the original position could not possibly occur. The original position is in fact a hypothetical construct from which one can derive principles of justice that are acceptable from a moral point of view. Because it plays a crucial role in the derivation of Rawls's two principles, its nature and function must be fully explored.

The original position is the "most philosophically favored interpretation of [the] initial choice situation for the purposes of a theory of justice" (p. 18). To satisfy ourselves that we have hit upon the most philosophically favored interpretation we must limit ourselves to *shared but weak conditions* and match the principles that would be chosen (given those conditions) against our considered moral convictions. Since we wish to derive moral principles from the original position, it is essential that the description of the original position be free of moral and value assumptions; otherwise the principles will only appeal to those already convinced. I shall follow Rawls in considering the original position under the headings of the circumstances of justice, the formal constraints of the concept of right, the veil of ignorance and the rationality of the contracting parties.

By the circumstances of justice Rawls means "the normal conditions under which human cooperation is both possible and necessary" (p. 126); in the main, his account follows that of Hume.<sup>5</sup> Society, Rawls maintains, is a cooperative venture for mutual advantage marked by both an identity and a divergence of interests; the identity makes social life possible and the divergence makes institutions necessary. Among the objective circumstances of justice are that individuals live in close

---

<sup>5</sup> See D. Hume, *A Treatise of Human Nature*, Book III, pt. II, § II (1740), and *An Enquiry Concerning the Principles of Morals*, § III, pt. I (1751).

proximity to one another, are vulnerable to attack and to having their interests blocked by others, are roughly similar in physical and mental powers and live under conditions of moderate scarcity. The subjective circumstances of justice include the fact that individuals, though capable of cooperation and a sense of justice, have collectively a variety of incompatible life-plans which demand conflicting shares of natural and social resources for their attainment. Further, men espouse incompatible philosophical, religious and political views. All of us, moreover, suffer the frailties of human nature: shortcomings of knowledge, thought, judgment and an assortment of moral faults.

Thus far, the description depicts life as it is. In rounding out his account, however, Rawls stipulates that the contracting parties are wholly self-interested; he does so in order "to insure that the principles of justice do not depend on strong assumptions" (p. 129). The fact that people form bonds of affection and take an altruistic interest in the interests of others can only strengthen adherence to principles adopted on the assumption of mutual self-interest.

Those in the original position think and act under certain constraints. Constraints of right hold for the choice of any moral principles, including principles of social justice. Rawls regards them as formal constraints, though he sees that they are not without moral force, since they rule out egoism. He emphasizes that "the conditions are not justified by definition or the analysis of concepts, but only by the reasonableness of the theory of which they are a part" (p. 131).<sup>6</sup> Just as theoretical entities in physics derive their meaning and justification from the theory of which they are a part, so too with Rawls's five formal constraints on the concept of right.

First, all principles must be general—avoid proper names and arbitrary definite descriptions. Second, all principles must be universal in application—they must hold for everyone, thereby ruling out self-contradictory and self-defeating principles. Third, all principles must be publicly chosen and acknowledged, thus disqualifying private, secret or otherwise esoteric moralities. Fourth, a conception of right must impose an ordering on conflicting claims, a task which Rawls admits is extremely difficult. And fifth, principles agreed to must be regarded as final and conclusive. Were we able to formulate a complete theory of morality, the totality of principles would not only determine what counts as a relevant consideration, but would assign weights as well. Conclusions from such principles would override self-interest, custom and law. A conception of right, then, "is a set of principles, general in form and universal in application, that is to be publicly recognized as a final

---

<sup>6</sup> In the main, Rawls follows the analysis of K. Baier, *The Moral Point of View*, ch. 8 (1958). There are differences, however. The most important is that Baier attempts to define the moral point of view from the very conception of morality; Baier's criteria define what it *means* to have a moral point of view. It is this attempt to avoid engaging in substantive, normative inquiry which Rawls rejects.

court of appeal for ordering the conflicting claims of moral persons" (p. 135).

Rawls introduces the notion of *pure procedural justice* to insure that any principles arrived at in the original position will be just. One adopts this procedure, he argues, when there is no independent criterion for the just outcome: "instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is, provided that the procedure has been properly followed" (p. 86). A distinctive feature of pure procedural justice is that the procedure must actually be used. In games of chance, any distribution of prizes is just, provided that fair procedures are actually used. It would be no good to hand out the prizes randomly before commencing play and then defend the distribution on the ground that actually playing the game according to purely procedural rules *might* have had exactly the same outcome. Although true, the defense is irrelevant.

Purely procedural justice will yield a just outcome, however, only if we assume that the contracting parties are shrouded in a *veil of ignorance*, a veil nullifying "the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage" (p. 136). Any information that would enable one to negotiate for principles tailor-made to one's own personal or historical circumstances is excluded. Thus, no one knows anything of one's natural abilities, place in society, aspirations, special psychological propensities, religious views or even the civilization to which one belongs (cf. p. 137). However, contractors in the original position are only selectively ignorant. Rawls assumes that they know a great many things most of us do not know: general facts about human society (*e.g.*, the circumstances of justice), the basis of social organization, the laws of psychology, the nature of politics, principles of economics and scientific theories generally (cf. pp. 137-38).

Particular information is restricted so that the parties have no basis for bargaining in the usual sense—for urging principles especially favorable to oneself. Since no one knows any distinguishing facts about oneself, there is no way to hold out for self-serving principles. General information is permissible, on the other hand, "since conceptions of justice must be adjusted to the characteristics of the systems of social cooperation which they are to regulate" (p. 138).

Individuals in the original position also know that they desire certain *primary goods*, *i.e.*, those goods normally necessary to or constitutive of anyone's life plan and therefore objects of everyone's rational desires. Natural primary goods, such as intelligence and health, concern us less than social primary goods, since only the latter are subject to distribution by social and political institutions. Among the most important social primary goods are rights, liberties, powers, opportunities, income, wealth and, above all, the bases of self-respect (cf. p. 62).

Rawls stipulates that individuals in the initial situation are rational, though rationality is construed narrowly as "taking the most effective



means to given ends" (p. 14). This account of rationality, standard in economic and social theory, entails that rational individuals have a coherent set of preferences, rank options according to purposes and adopt plans satisfying more, rather than fewer, desires (cf. p. 143). By eliminating moral preferences from his account of rationality, Rawls believes that principles derived from his amoral construct are more likely to be accepted. To be sure, Rawls builds into the original position what he calls "a thin theory of the good" in order "to explain the rational preference for primary goods and to explicate the notion of rationality underlying the choice of principles in the original position" (p. 397). A thin theory of good "is necessary to support the requisite premises from which the principles of justice are derived" (*id.*). Because the thin theory is confined to bare essentials that no one can deny, principles derived from these essentials are not in danger. Rawls does depart from the standard economic account of rationality in one respect: he assumes that individuals in the original position, or Rational Individuals as I shall refer to them, are not envious—Individuals who have gained a favorable position at no cost to other Individuals will not be viewed with hostility. Were envy permitted, the original position would yield unacceptably egalitarian principles.

Explicit in Rawls's account of rationality is the belief that Individuals have ends, though not that mankind has any specific end. Pursuit of *any* end must be regulated by those principles accepted in the original position on the assumption that Individuals are wholly self-interested and have a variety of final ends. This is part of what Rawls means when he says, "in justice as fairness the concept of right is prior to that of the good" (p. 396). To use standard jargon, justice as fairness is a deontological, not a teleological, theory.<sup>7</sup> Rawls concludes that those principles which Rational Individuals *would* accept in the original position *are* the correct principles of social justice that we, as flesh and blood individuals, *should* apply to the basic structure of our (or any) society.

Professor Rawls's characterization of the original position is troubling. Only if Individuals are rational do their negotiations and principles merit our attention, let alone our acceptance. To avoid introducing into the concept of rationality any controversial ethical statements which would jeopardize wide acceptance of the original position as a satisfactory analytic construct, rationality is interpreted as taking the most effective means to given ends. Neither the interpretation nor its rationale are sound.

Throughout *A Theory of Justice* Rawls assumes that if everyone would agree to an interpretation or principle when veiled in ignorance,

---

<sup>7</sup> Teleological theories assert that the only consideration which makes an action or rule right or obligatory is the goodness of its consequences; deontological theories claim that certain features of the act itself other than the value it brings about must be considered. For a clear, brief introduction to ethics and kinds of ethical theories, see W. Frankena, *Ethics* (1963).

then once that veil lifts, one can only supplement, but never radically alter or replace, that interpretation or principle. Principles of risk accepted in the original position, for instance, are carried over intact to situations where one does not labor under full ignorance. Yet surely what might be agreed to under Rawls's special, stipulated conditions would likely be unconvincing or even irrational under normal conditions; policies which are reasonable under conditions of limited knowledge may well be unreasonable given substantially more knowledge. By holding Individuals to what they agree to when veiled in ignorance, Rawls no doubt avoids messy disputes regarding the relevance and weight of various considerations—but only at the cost of distorting the concept of rationality.

Rationality interpreted as efficiency may sound plausible when ends are not in dispute. Unfortunately, this interpretation glosses over the fact that people have a multiplicity of ends ordered and related in complex and incompatible ways. Nor can a sharp distinction be drawn between means and ends; means and ends are so related that, except in the simplest cases, an alteration of one affects the nature of the other. One cannot suppose, for example, that every effective means of teaching people always results in their having the same education; a student who studies Shakespeare by means of a programmed text is likely to have a different understanding and appreciation of the subject than one who learns Shakespeare through the theater, though both students may possess the same facts.

Rawls's interpretation of rationality is defective in another way. Many people value certain ends above any purely formal procedure of decisionmaking, yet Rawls would have them sacrifice these ends for a procedure they would have adopted had they been Rawls's Rational Individuals. If there is an end that a man is *convinced* (rightly or wrongly) requires the complete devotion of everyone, he would be unlikely to limit his pursuit of this end whenever it ran afoul with a procedure he did not in fact adopt. There seems to be no procedural shortcut which will eliminate conflicts over ends or, indeed, entire ways of life. Rawls, we will recall, contends that purely procedural justice is rational "when there is no independent criterion for the right result" (p. 86). Perhaps, but this will not persuade those who believe that such a criterion exists and, what is more, that they know what it is. That there is great disagreement concerning the nature of this criterion is neither here nor there; a person who is convinced that a certain arrangement is just will not usually be moved to adopt a principle of pure procedural justice merely because others disagree with him. Rational Individuals adopt a purely procedural principle because they have no deep moral and value convictions; real people who have such convictions will resist suppositions which require that they set these convictions aside. Rawls forbids his Rational Individuals to have a full theory of the good because he does not wish to jeopardize the prior

place of the concept of right, but this unfairly tips the scale against teleological views at the outset.

The very intelligibility of the original position is not beyond question. Can Individuals be unaware of the civilization to which they belong, yet cognizant of the laws of economics, the principles of psychological development and the nature of political affairs? Because physical laws are unaffected by cultural factors, may we assume that the same is true of psychological, sociological and economic laws? Once Rawls spells out the nature of these laws in detail, as he must, one may find that they are only applicable to some societies at a certain stage of their historical development. Further, if man is necessarily a social being, *i.e.*, a creature whose nature cannot be adequately understood apart from his social relationships, then to develop a theory of justice from the viewpoint of theoretically defined individuals living in splendid isolation from any particular social and historical context is misguided. If Jeremy Bentham took human nature to be that of the English shopkeeper, as Karl Marx alleged, at least he took the point of view of actual people situated in an actual historical situation; Rawls's Rational Individuals, on the other hand, correspond to no one in any social or historical situation.

Behind these specific criticisms lies a nagging feeling that the original position merely reflects, in an abstract way, contemporary liberal democratic attitudes—the description of it has appeal precisely because it sums up our liberal, individualistic outlook. Perhaps this is inevitable; given Rawls's pragmatic approach we are bound to find a prominent place for our firmly held moral beliefs. Yet unless our attitudes and beliefs are grounded in more than consensus, their appeal will be less widespread than Rawls anticipates. My concern, to put the point differently, is that any principles derived from the original position will simply reproduce *in theory* what we heirs to the liberal, democratic Western tradition intuitively believe *in fact*. Unless we are provided with the metaphysics of a theory of justice, which Rawls studiously avoids, I do not see how Rawls can overcome this essentially Hegelian objection.

Professor Rawls's decision to cast his theory in terms of an ideal agreement or social contract is also troublesome. Specifically, Rawls never adequately describes the tasks the social contract model performs or how literally the model should be taken. Rawls clearly believes that the concept of social justice can be illuminatingly analyzed in terms of a contract which Rational Individuals in the original position would unanimously make. Precisely what is shown, however, is never made clear. In particular, it is not clear whether the original position and the social contract are intended as devices (1) for *analyzing* a conception of justice we already have, (2) for *generating* principles of justice which then must be justified in noncontractarian ways or (3) for *justifying* as well as *generating* principles of social justice.

I am inclined to think that Rawls uses the original position and the social contract simply as tools of analysis and devices for generating principles of justice, and not as a way of justifying these principles. I do so not because the weight of textual evidence clearly points in this direction, but because it seems obvious that hypothetical contracts agreed to by hypothetical individuals in a hypothetical situation cannot bind me. How could they, when what I myself *would* have done (but did not) cannot bind me. *Had* a particular student asked me to direct his research, I might well have promised to do so; given that I made no such promise, however, I am now under no obligation to him. Therefore, if I am not bound by a hypothetical contract that I myself would have made, I am surely not bound by a hypothetical contract that certain mythical creatures would make.

On this interpretation of Rawls, the social contract functions as a way of perceiving, conceiving and construing that part of morality concerned with justice. Although other perspectives abound, none is as fruitful. The concepts of equality and the fortuitousness of human differences in nature and situation are among the central notions Rawls's account attempts to explain; these notions are also among our deepest moral intuitions. The social contract itself may be an artificial notion. But, as David A.J. Richards, a proponent of Rawls's method, declares,

there is nothing artificial or superficial about the ideas which the ideal contract view is trying to articulate: these ideas are as real as our oldest and most haunting collective myths. The point of the theoretical elaborateness of the ideal contract view is not artificially to sophisticate these profound and durable ideas, but rather to provide or suggest some way in which these ideas may be so expressed that one may fruitfully indicate how these notions establish the specific and determinate structure of our moral experience. The notion of contractors deciding what count as moral principles must be understood from this point of view: as a way of giving workable expression to certain deep and permanent features of moral thought. . . .<sup>8</sup>

Like any other perspective, that of the social contract must be judged by its fruits. To Richards this means that the social contract model must ultimately be judged

in terms of the creative usefulness of such theoretical intricacy in providing tools with which we may take hold and better understand not only the nature of some of the most abstract and difficult concepts which we employ, but also how such concepts are specifically applied to the variegated circumstances of human life.<sup>9</sup>

I have spoken of the social contract as a *model* from which one can rigorously deduce principles of justice, and certainly there are passages to justify this interpretation. Rawls claims, for example, that his argument "aims eventually to be strictly deductive. . . . We should strive

---

<sup>8</sup> D. Richards, *A Theory of Reasons for Action* 91 (1971). Rawls supervised Richards's undergraduate senior thesis, and Richards acknowledges the considerable influence Rawls's own work has had on him; clearly, the influence is enormous.

<sup>9</sup> *Id.*

for a kind of moral geometry with all the rigor which this name connotes" (p. 121). Other passages suggest that the social contract and the original position are merely vivid *metaphors* which play no essential role in the argument. Thus, we find Rawls asserting:

To say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion. *If necessary, the argument to this result could be set out more formally. I shall, however, speak throughout in terms of the notion of the original position.* It is more economical and suggestive, and brings out certain essential features that otherwise one might easily overlook (p. 138 (emphasis added); cf. pp. 21-22).

Whether model or metaphor, the social contract and the original position give rise to certain misgivings as regards their usefulness. Some of these have already been discussed; others will be mentioned as we examine Rawls's principles of justice.

### III

#### TWO PRINCIPLES OF SOCIAL JUSTICE

Rational Individuals in the original position, we are told, would adopt two principles of justice ranked in a certain order. According to the *principle of equal liberty*, "[e]ach person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all" (p. 302). This principle, Rawls argues, takes precedence over the principle regulating economic and social inequalities, for, assuming a minimum level of general well-being, it is "irrational from the standpoint of the original position to acknowledge a lesser liberty for the sake of greater material means and amenities of office" (p. 542). Economic and social benefits (and burdens) are to be distributed in accordance with the *principle of equal benefits and opportunities*: "Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged . . . and (b) attached to offices and positions open to all under conditions of fair equality of opportunity" (p. 302).

Assuming a minimum level of general well-being, liberty can be restricted only if a less extensive liberty strengthens the total shared system of liberty or if unequal liberty is acceptable to those who will have less liberty. Liberty of conscience and the rights which define the integrity of the person are our fundamental basic liberties; certain political rights, on the other hand, may be foregone either to strengthen the entire system of justice or to compensate those less fortunate (cf. p. 247). Yet even these inequalities are justified only as long as social conditions are being developed under which less than equal liberty would be unjustified. Rawls's purpose in ordering the two principles as he does is to prohibit exchanges between basic liberties and economic or social benefits. *If* minimum human wants are satisfied, economic considerations cannot justify suspension of any basic liberties, including

the right to vote and to hold public office, freedom of speech and assembly and the right to hold personal (not necessarily private) property.

Once the principle of equal liberty is satisfied, one might be tempted to adopt the principle of Pareto optimality to determine a just distribution of goods. According to this principle, "a configuration is efficient whenever it is impossible to change it so as to make some persons (at least one) better off without at the same time making other persons (at least one) worse off" (p. 67). Although Rational Individuals would accept this principle to judge the efficiency of their economic and social arrangements, Rawls argues convincingly that they would reject it as a principle of justice, since there are usually many arrangements which, though efficient, are manifestly unjust—including the limiting case where one person (or class) receives every available benefit.

If Rational Individuals would reject the principle of efficiency, so would they also reject any utilitarian principle. Classical utilitarianism, which bids us to maximize total (net) benefits, permits and even requires some people to carry enormous burdens provided only that their sacrifices maximize utility over the whole society. Since we are assuming that utility cannot serve as the first principle of justice and that the principle of equal liberty, which is the first principle, has been satisfied, the worst cases of injustice could not occur. Even so, there would seem to be no reason why a Rational Individual would choose a principle of distribution which might well be to his disadvantage however much it might benefit society as a whole. For similar reasons, Rational Individuals would reject suggestions that average (net) utility be maximized.

Rational Individuals in the original position, Rawls argues, would adopt the *difference* principle. Of the various efficient arrangements, the difference principle singles out that distribution of benefits which improves the expectations of the least advantaged members of society:

Assuming the framework of institutions required by equal liberty and fair equality of opportunity, the higher expectations of those better situated are just if and only if they work as part of a scheme which improves the expectations of the least advantaged members of society. (p. 75).

Admittedly, not all arrangements satisfying this principle will maximize total or even average utility. However, Rational Individuals would choose a principle promising less than the utility principle because, as rational creatures, they would adopt a *maximin policy*; that is, a policy of accepting only those principles which have the best worst result. People who wear raincoats on cloudy days act on this policy—though a nuisance, a raincoat guarantees the best result (keeping dry) should the worst conditions obtain (rain), though at the cost of forfeiting one's chances for the best possible outcome (neither rain nor raincoat). As Rational Individuals have no idea of their eventual position in society, they would subscribe to a maximin rule. This rule, though not part of Rawls's definition of rationality, is presumed to guide the choice of both principles of justice.

By distinguishing between equal liberty and opportunity on the one hand and the *worth* of equal liberty and opportunity on the other hand, Rawls acknowledges that poverty, ignorance and natural liabilities may render formal equality virtually worthless. However, I am not sure that Rawls traces the implications of this distinction correctly. If everyone has equal liberty, the question of compensation for lesser liberty obviously does not arise; but how are we to compensate the disadvantaged for the lesser *worth* of liberty? "The lesser worth of liberty is . . . compensated for, since the capacity of the less fortunate members of society to achieve their aims would be even less were they not to accept the existing inequalities whenever the difference principle is satisfied" (p. 204). This is true but irrelevant. The disadvantaged would indeed be ill-advised to take less than the difference principle authorizes *given that equality of liberty must be maintained*. But this begs the question. The disadvantaged might argue that, because their basic liberties are worth so little, they ought to be given *greater* liberties and rights than the advantaged enjoy, e.g., double votes. Just this sort of argument has in fact been advanced by radical blacks in an effort to secure automatic representation on important decisionmaking bodies. The argument may be unsound, unjust or both, but Rawls does not show that it is. One might even argue that Rational Individuals obeying a maximin rule would agree to inequalities of basic liberties in favor of disadvantaged groups whenever the worth of liberty falls below a certain level or whenever certain human needs are being ignored by existing social institutions; after all, such inequalities would seem to maximize one's well-being were one to end up among the least advantaged.

The maximin rule itself needs clarification and support. In explaining the rule, Rawls says that "the two principles are those a person would choose for the design of a society in which his enemy is to assign him his place" (p. 152). Were one's position indeed determined by one's enemies, Rawls's defense of the principles makes sense. However, Brian Barry points out that "our hypothetical principle-choosers are *not* going to be assigned their positions by their enemy."<sup>10</sup> Rather, their positions will depend on personal characteristics and good fortune. This play-safe strategy is therefore unnecessary.

Consider again the man who carries a raincoat on cloudy days. If he believes that there exists a malevolent being intent on causing him to suffer, then of course the maximin rule is reasonable; if he believes on the other hand that there exists a benevolent being intent on causing him to prosper, then a maximax rule, prescribing that he opt for the best possible outcome, is reasonable. But supposing, as he surely would, that the weather is not affected by his deliberations, our rational man would estimate the likelihood of rain and dress accordingly. Individuals pursuing their own interests exclusively might choose those principles leading to a high average level of well-being; "whether they would prefer one with more equal or less equal dis-

---

<sup>10</sup> Barry, *On Social Justice*, Oxford Rev. 36 (Trinity Term 1967).

tribution," Barry concludes, "would depend on their taste for gambling."<sup>11</sup>

Rawls introduces the veil of ignorance just to meet such objections; under conditions of ignorance, insists Rawls, Rational Individuals could not play the percentages and would therefore adopt the maximin rule. But would they? Wouldn't it be at least as rational to agree to a social minimum and then devise basic institutions which would permit extreme inequalities? For no matter where one ended up in society, one would have a roof over one's head and other necessary creature comforts. Once this floor had been decided upon, however, Rational Individuals might choose to gamble for stakes involving substantial inequalities, especially if the inequalities were thought to be necessary to purchase a high level of art, learning or even a hedonic paradise.

However, supposing that Rational Individuals would subscribe to a maximin rule, Rawls stops short of its full implications. He assumes without argument that Rational Individuals would choose principles applicable only to *representative* persons, not to *particular* or *actual* persons. Yet if Rational Individuals are completely self-interested, Rawls cannot suppose that they would be concerned with the interests of any representative or typical or average person of any group; for no one knows whether he will be representative, typical or average. Thus the expectations of representative persons are irrelevant. To use Rawls's own heuristic device, imagine that one's enemy were to assign one's place in society. Adopting the maximin rule, everyone would choose only those principles guaranteeing that any economic or social distribution must improve the expectations of *each and every* member of the least-advantaged group. Rawls's attempt to avoid the strict egalitarian implications of the maximin principle therefore fails.

These criticisms make apparent that the social contract conception of justice can generate any number of principles depending on the conditions built into the interpretation of the original negotiating position. Why should we accept the conditions Rawls imposes on the initial choice situation? Rawls addresses this question twice, once at the beginning of *A Theory of Justice* and once at the very end; both times the answer is the same. Despite the hypothetical nature of the original position and of the social contract,

the conditions embodied in the description of the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded to do so by philosophical reflection. Each aspect of the contractual situation can be given supporting grounds. . . . (p. 21; see p. 587).

"Philosophical reflection" appears to refer to Rawls's belief that moral philosophy is Socratic:

[W]e may want to change our present considered judgments once their

---

<sup>11</sup> Id. at 38.



regulative principles are brought to light. And we may want to do this even though these principles are a perfect fit. A knowledge of these principles may suggest further reflections that lead us to revise our judgments. . . . (p. 49).

The resulting reflective equilibrium is therefore much more than matching one's existing judgments with moral principles; the kind of reflective equilibrium which concerns moral philosophy involves an investigation of various conceptions of justice and a consideration of philosophically relevant arguments on behalf of each. The original position and the social contract, therefore, are expository devices which draw attention to conditions we either accept or can be brought to accept.

Rawls stresses that his theory of justice is a theory of the moral sentiments; that is, it sets out the principles governing our moral capacities and powers, especially our sense of justice. Limitations of space prevent a detailed exposition and evaluation of this aspect of Rawls's theory. I mention it, however, because of its bearing on the question of justification of the two principles of justice. We have already seen that definitions and analyses of meaning have no special place in Rawls's approach; they stand or fall with the theory itself. The same is true, or is suggested to be true, of justification. "[I]f we can find an accurate account of our moral conceptions," Rawls declares, "then questions of meaning and justification may prove much easier to answer. Indeed," he concludes, "some of them may no longer be real questions at all" (p. 51). Just as developments since Frege and Cantor have immeasurably deepened our understanding of the meaning and justification of statements in logic and mathematics, so Rawls hopes that if he can provide an adequate theory of our moral sentiments and concepts, a similar transformation may occur in ethics.

#### IV

#### CONCLUSION

Readers of this journal will find that *A Theory of Justice* illuminates a variety of topics and controversies I have slighted or ignored altogether: the nature and justification of conscientious refusal, civil disobedience and political obligation; the rule of law and the status of majority rule; the limits of tolerance and conscience; guilt, shame and happiness; duties, natural duties and obligation; and the nature of a well-governed society. Rawls's analyses are sensible, subtle and occasionally yield surprising conclusions. He suggests, for instance, that citizens generally (as opposed to officeholders) have no political obligation (see p. 114) and that the good things of life should *not* be distributed according to moral desert (see p. 310).

There is much for everyone to learn from this book, though each reader will have his own reservations and concerns. But this much is beyond doubt: *A Theory of Justice* makes an outstanding contribu-

tion to moral, social and political thought; it deserves its inevitable fate of being the most discussed book in its field for the next decade.

To see our place in society from the perspective of contractors in the original position, Rawls concludes, is to see it *sub specie aeternitatis*: "it is to regard the human situation not only from all social but also from all temporal points of view" (p. 587). If *A Theory of Justice* actually provides us with a temporal perspective tied to liberal, democratic preconceptions, a perspective which may already be on the wane, we may well recall Hegel's observation that the owl of Minerva spreads its wings only at dusk.

HANS OBERDIEK\*

---

\* Professor of Philosophy, Swarthmore College.