

CHAPTER IV

RAWLS'S THEORY OF JUSTICE: A SUMMARY

Like most systematic philosophers, John Rawls insists that the parts of his theory can be properly understood and validated only in the context of the coherent whole (for example, see §9 and §87 of *A Theory of Justice*). Among philosophers this is a common, and all too often reflexive response to valid criticism. In Rawls's case, however, the plea is well taken. His book contains a marvelously integrated system of mutually reinforcing parts, and the thoughtful reader is inclined to conclude, with Robert Nozick, that "it is impossible to finish [Rawls's] book without a new and inspiring vision of what a moral theory may attempt to do and unite, of how *beautiful* a whole theory can be" (1974, p. 183).

Accordingly, the systematic integrity of Rawls's theory of justice requires a presentation of a general account of this theory. Unless the reader understands Rawls's basic theory, his position concerning "justice between generations" will be incomprehensible, and a critical examination of this position will be unenlightening. The purpose of this chapter, then, is simply to make the remainder of the dissertation intelligible to the reader who has not carefully read Rawls's *A Theory of Justice*. (The reader who is familiar with Rawls's theory may therefore elect to omit this chapter and to resume with Chapter V.) Once the basic theoretical substratum has been set down in this chapter, we will turn, in the next, to an exposition of Rawls's explicit views concerning *duties, rights, and justice to posterity*. In the concluding chapters, Rawls's position concerning posterity will be critically examined.

In the brief span of this exposition it will, of course, be impossible to convey the rich detail, broad scope, and elegant structural coherence of Rawls's book. Moreover, I will have little opportunity to present, and still less to examine, Rawls's numerous supporting arguments or the many critical responses that have followed the publication of the book.¹ This chapter has been designed to serve the purposes of this dissertation. Rawls's broader purposes are best set forth and advocated by Rawls himself in his profound, provocative, and memorable book.

20. "Justice as Fairness": A Preliminary Assessment

Rawls's task: an overview. "Justice as Fairness," as Rawls prefers to call his theory of justice, stands apart from the work of contemporary moral philosophers. Unlike the logical positivists, and many of their analytical successors, Rawls believes that moral discourse has an objective, rational, and cognitive basis. Ethical controversy, he claims, may be more than a conflict of emotions, attitudes, or prescriptions. In fact, unlike many contemporary moral philosophers, Rawls is not simply offering "a description of ordinary meanings" (Rawls, p. xi), albeit his method and style have been favorably influenced by his familiarity with contemporary trends and techniques in conceptual analysis. Finally, Rawls is not content to confine himself to "meta-ethical" descriptions of moral discourse and

controversy, though his frequent second-order reflections upon the meaning and method of his developing system are quite illuminating.

Contrary to the prevailing fashions of contemporary moral philosophy, Rawls has prepared a careful, coherent, empirically supported theory of *substantive, normative*, ethics (p. ix). In other words, Rawls's primary purpose is not merely to describe what philosophers, or ordinary users of the English language, are in fact doing when they use moral discourse or what they mean when they utilize such words as "right," "duty," "good," "virtue," or "justice." His purpose, like that of Plato and Aristotle of old, is to discover and to determine substantive answers to such perennial moral questions as: "By what principles should I direct my life?" "What should I live for?" "What is a just community?" "Why should I be moral?" Thus, while "Justice as Fairness" is a departure from the predominant fashions of contemporary moral philosophy, it is at the same time a return to issues that have been conspicuous in the larger portion of the history of ethical thought.

Rawls's quest for the universal principles of justice involves several subsidiary tasks. Prominent among these: (a) he wishes to identify and explicate the general moral principles which seem to underlie "well-considered" moral judgments in ordinary life. (The reader's task will be facilitated if he keeps in mind that "judgments" usually denotes particular, practical moral decisions, and "principles" refers to abstract and general moral rules.) (b) He seeks to determine whether these ordinary principles would be agreed to under fair conditions of objective deliberation and reflection (hence, the term "justice as fairness"). (c) He wishes to demonstrate that acceptable "fair" principles are practically workable and would lead to a just, stable, "well-ordered" social arrangement in light of the information available from the social sciences (Daniels, 1975, p. xiv). Finally (d), Rawls wishes to determine whether such a well-ordered ("just") society is "congruent" with individual human good; that is, whether a just society would support and be supported by individuals whose personal lives were conducted according to "rational Life plans." Throughout his book, Rawls's prevailing approach is *deontological, objective, socially oriented*, and in the tradition of *contract* theory. I will next examine these key concepts in order.

Justice as fairness is a deontological theory. The distinction between the *Right* and the *Good* is basic to Rawls's theory. He interprets these crucial terms as follows: *The Right* applies to actions and circumstances in accordance with principles "which rational persons concerned to advance their interests would accept in [a] position of equality to settle the basic terms of their association" (pp. 118-119). By *The Good*, Rawls understands that if "an object has the properties that it is rational for someone with a rational plan of life to want, then . . . it is good for him. And if certain sorts of things satisfy this condition for persons generally, then these things are human goods" (p. 399). (For a fuller development of the concept of Good, see Rawls, §61.) In the course of this chapter and beyond, both of these concepts will undergo considerable elaboration and refinement.

It is important to note that Rawls, as a deontologist, holds that the *Good* is subordinate to the *Right*, and thus, that the maximization of the Good is not sufficient to validate the rightness of an act or a rule. Rawls is clear and unequivocal about this:

The principles of right, and so of justice, put limits on which satisfactions have value; they impose restrictions on what are reasonable conceptions of one's good. In drawing up plans and in deciding on aspirations men are to take these constraints into account. Hence in justice as fairness one does not take men's propensities and inclinations as given, whatever they are, and then seek the best way to fulfill them. Rather, their desires and aspirations are restricted from the outset by the principles of justice which specify the boundaries that men's systems of ends must respect. We can express this by saying that in justice as fairness the concept of right is prior to that of the good. A just social system defines the scope within which individuals must develop their aims, and it provides a framework of rights and opportunities and the means of satisfaction within and by the use of which these ends may be equitably pursued. The priority of justice is accounted for, in part, by holding that the interests requiring the violation of justice have no value. Having no merit in the first place, they cannot override its claims. (p. 31)

The deontological standpoint is fundamental to Justice as Fairness, and sets it irreconcilably apart from its significant rivals, utilitarianism and perfectionism. (For more on Rawls's distinction between *the right* and *the good*, see §26, below).

Reason and experience in moral inquiry. Unlike such "emotivists" as A. J. Ayer and C. L. Stevenson, not to mention many other "non-cognitivists," Rawls believes that moral assertions have cognitive meaning, that objectively sound arguments can be presented in behalf of certain moral beliefs, and that objective empirical facts can be cited in support of such assertions and beliefs.² Furthermore, as Stuart Hampshire observes in his highly favorable review, Rawls believes that moral judgments can be "shown to be instances of a few general principles at work," and thus, we can "have an assurance that our moral beliefs have a rational foundation" (1972, p. 34). Thus, both *reason* and *experience* play important parts in effective moral deliberation. Rawls, of course, does not claim that ethics is just another social science. He does not, that is to say, deny the basic philosophical tenet that there are fundamental differences *in kind* between assertions of fact and value judgments. He does, however, deny that the distinction is so complete that statements of fact have no relevance to moral discourse. On the contrary, he draws heavily from such diverse fields as developmental psychology, social psychology, economics, and political science. Conversely, Rawls's moral theory bears significant implications for these disciplines, as is evident in the numerous responses to Rawls's theory that continue to appear in social, economic, political, and legal journals.

It would be a mistake, however, to overstate Rawls's position concerning the objectivity of moral judgments. In Justice as Fairness, moral insight is not entirely a matter of discovering immutable facts of human nature, or of drawing logically pure deductions from clear, distinct and eternal *a priori* principles. In the first place, our moral judgment is dependent upon changing conceptions of human need and desire. But even more, in moral deliberation there is a need for creative intelligence, not unlike that of the theoretical scientists, as the moral agent seeks to connect disparate moral judgments, motives, and aims under comprehensive and abstract principles of justice. Hampshire (1972) aptly summarizes the inalienably human element in Rawls's system, as he notes that:

[Justice as Fairness] opposes the designed and invented moral order to the blind causality of the natural order, the moralized and socialized citizen to the natural man. Moral custom and rule are not second nature, but human artifice Equal liberty, and the requirement that each man counts for on in reckoning the common good, are in practice recent inventions, principles of 1789. (p. 38)

An emphasis upon social justice. "The primary subject of justice," writes Rawls, "Is the basic structure of society" (p. 7). Rawls's concern with the social dimension of human behavior and evaluation pervades his long book. In the following significant passage early in the book, Rawls presents his concept of society and indicates that social behavior necessarily involves both *cooperation* and *conflict* among individuals within the community. From these fundamental features of social activity, he argues, arises the need to establish shared rules of conduct; that is to say, *principles of justice*. Says Rawls:

Let us assume, to fix ideas, that a society is a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them. Suppose further that these rules specify a system of cooperation designed to advance the good of those taking part in it. Then, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts. There is a conflict of interests since persons are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. A set of principles is required for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These principles are the principles of social justice: they provide a way of assigning rights and duties in the basic institutions of society and they define the appropriate distribution of the benefits and burdens of social cooperation. (p. 4)

For all his interest in the social aspects of justice, Rawls's concern for the rights of, and fairness to, the individual is paramount. *Institutions*, he insists, must be designed to fit the just needs and rights of the individual. Furthermore, the rights and just interests of the few are never to be compromised for the sake of the greater benefits of the many. Thus, Rawls endorses a common and basic deontological criticism of utilitarianism. Finally, as we shall see, Rawls derives his fundamental principles of justice through a carefully devised thought-experiment which postulates the rational moral decisions that might be reached by an assembly of self-serving, prudent *individuals*.

The limits of justice as fairness. Rawls's theory of Justice, and his parallel theory of the Good (developed in the final third of the book), cover a large portion of the traditional domain of moral philosophy. However, as Rawls readily acknowledges, some significant areas of moral concern are left out. In particular, he acknowledges that Justice as Fairness "would seem to include our relations with other persons and to leave out of account how we are to conduct ourselves toward animals and

the rest of nature" (p. 17). Later, he suggests that "a correct conception of our relations to animals and nature would seem to depend upon a theory of the natural order and our place in it" (p. 512). However, such a theory is not attempted in *A Theory of Justice* and, considering both the scope and the depth that he otherwise achieves, perhaps Rawls should not be faulted for this limitation. Even so, later in this dissertation I will suggest that man's place in nature, and his conception thereof, bear profound implications for the question of the duty to posterity (§42, below).

21. Contract Theory

As early as the third paragraph of his preface to *A Theory of Justice*, Rawls asserts that he has attempted "to generalize and carry to a higher order of abstraction the traditional theory of the social contract as represented by Locke, Rousseau, and Kant" (p. viii). Rawls understands the central problem of the "social contract" to be the derivation and articulation of principles of justice that rational persons, cognizant of all relevant information and free of circumstantial bias, would accept as fair procedures of social order--procedures, that is, to which they would willingly submit. As he develops his theory, Rawls is mindful of the pitfalls that have plagued deontological theories in the past. Conspicuous among these difficulties are (a) the proliferation of distinct moral precepts without a coherent structure among these precepts, or a ranking by priority, and (b) a failure to justify these precepts with grounds more substantial than "intuition" or "common sense" (Schwartz, 1973, p. 295). Rawls believes that his contractarian approach to justice provides both a workable set of priority rules and a justification procedure based upon objective principles of rational choice.

"Rational choice." The notion of "rational choice" is integral to Rawls's characterization of his theory as "contractarian." As Rawls expresses it:

The merit of the contract terminology is that it conveys the idea that principles of justice may be conceived as principles that would be chosen by rational persons, and that in this way conceptions of justice may be explained and justified. The theory of justice is a part, perhaps the most significant part, of the theory of rational choice. (p. 26)

In the following crucial passage, Rawls further describes these principles of justice as:

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. (p. 11)

We have now reached the threshold of the central conceptual model of Rawls's system – an hypothetical assembly of "rational contractors" which Rawls chooses to call "the original position."³

We are to 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. Men are to decide in advance how they are to regulate their claims against one

another and what is to be the foundation charter of their society A group of persons must decide once and for all what is to count among them as just and unjust. (pp. 11-12)

Rawls, of course, does not suggest that the original position describes an historical event, much less a primitive state of culture. It is, he says, "a purely hypothetical situation characterized so as to lead to a certain conception of justice" (p. 12) Even so, the original position describes a perspective that we can simulate in our own moral deliberations (p. 120. See also p. 19). Indeed, *A Theory of Justice* represents Rawls's comprehensive attempt to do just that.

An ideal, "strict compliance" theory. A society that is "effectively regulated by a public conception of justice" and "designed to advance the good of its members" is said, by Rawls, to be "well ordered." It is "a society in which everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions generally satisfy and are generally known to satisfy these principles" (p. 5). Rawls is primarily interested in the principles that would regulate such a society. He is, that is to say, developing what he calls a "strict compliance" rather than a "partial compliance" theory. "The reason for beginning with an ideal theory," he says, "is that it provides, I believe, the only basis for the systematic grasp" of such pressing "non-ideal" problems as punishment, the just war, civil disobedience, and other means of dealing with injustice (pp. 8-9). His strategy, then, is to derive principles that apply to the ideal "well ordered society" under favorable conditions, and then to evaluate, from this perspective, the all too commonplace departures from this ideal that are to be found in the actual world. "Viewing the theory of justice as a whole," says Rawls:

The ideal part presents a conception of a just society that we are to achieve if we can. Existing institutions are to be judged in the light of this conception and held to be unjust to the extent that they depart from it without sufficient reason Thus as far as circumstances permit, we have a natural duty to remove any injustices, beginning with the most grievous as identified by the extent of the deviation from perfect justice. (p. 246)

Why by moral? In Plato's *Republic*, Glaucon relates the fable of Gyges who, through the possession of a magical ring, was able to make himself invisible and thus gain the power to commit crimes and injustices without fear of detection and punishment. Upon relating this tale, Glaucon posed a challenge which has persisted throughout the centuries: Why should Gyges, or anyone else immune from external sanctions, rebuke, and punishment, submit to rules of justice? What advantage remains, to the agent himself, of being moral? The brief glimpse that we have had so far of Rawls's theory suggests the contractarian response: acceptance of the rules of justice by *all* members of a community provides personal advantages to *each* individual member. Says Rawls:

A common conception of justice and the public awareness of men's willingness to act in accordance with it, are a great collective asset Having trust and confidence in one another, men can use their public acceptance of these principles enormously to extend the scope and value of mutually advantageous schemes of cooperation. (pp. 347-348)

"Just a moment," the Nietzschean aristocrat would retort, "this is nothing but a slave morality: You are attempting here to foist constraints upon the strong and the noble in the interests of the petty, the weak and the fearful. Why should an outstanding and self-sufficient person submit to such unworthy trifles?" The reason, quite simply, is that without a public acknowledgment and observance of rules of right conduct, *all* are weak in the face of the capricious forces of nature, widespread conniving and conspiring for personal security and advantage, and the consequent Hobbesian "warfare of all against all." In words that Rawls would readily endorse, Michael Scriven (1966) observes that:

Each citizen's chances of a satisfying life for himself are increased by a process of conditioning all citizens *not* to treat their own satisfaction as the most important goal. Specifically, a system which inculcates genuine concern for the welfare of others is, it will be argued, the most effective system for increasing the welfare of each individual. Put paradoxically, there *are* circumstances in which one can give a selfish justification for unselfishness. (p. 240)

And there is, to Rawls, a further paradox: by relinquishing his "freedom" to violate just precepts at convenience and whim, the individual gains the far more valuable freedom to pursue, with confidence, his own personal good – that is to say, his own "rational plan of life."

22. "Reflective Equilibrium"

To characterize Justice as Fairness as a "contract theory" both tells us a great deal about this theory and leaves a great deal unexplained. There are, after all, numerous types of contract theory, each evolving from a differing conception of the nature of the contracting parties, their motives, intentions, and aims, the knowledge permitted or excluded from their deliberations, the rules of evidence and other formal factors in the decision-making, and so on. Each of these considerations will bear upon the ultimate decisions of the contractors. In other words, in contract theory, the principles of justice follow from the characterization of the hypothetical original assembly of contractors. Thus, the crucial question is: "according to what criteria are we to select from the available contractual models?"

Rawls's partial answer to this fundamental question is quite straightforward: that contractual arrangement is to be preferred which entails principles of justice that are consistent with our "moral sense," and with the "considered moral judgments" that we make in the course of ordinary life. Thus, he writes quite candidly:

What counts is whether the conception of justice as fairness, better than any other theory presently known to us, turns out to lead to true interpretations of our considered judgements, and provides a mode of expression for what we want to affirm. (p. 452)

But if the test of moral principles is to be an agreement with ordinary moral judgments, why then bother with the elaborate machinery of a hypothetical contract? Doesn't the theory thus reduce to mere *intuitionism*, or worse, *conventionalism*?

Rawls believes that he can acquit his theory of both charges. He believes, let us recall, that *partial* validation of the theory is to be found in the *considered* moral judgments of ordinary life. The crux of his defense against the charges of *intuitionism*, *subjectivism*, and *conventionalism*, is to be found in these key words "*partial*" (validation) and "*considered*" (moral judgments). We will soon encounter additional criteria for judging competing models of the contractual assembly. However, let us turn immediately to the question of "considered moral judgments" in ordinary life.

"*Considered moral judgments.*" Rawls's interest in everyday moral judgments is manifested in three rather commonplace observations. *First*, he notes that virtually all people everywhere do, in fact, make moral judgments. (No matter, for the moment, that the content or the *priority order* of these judgments might differ.) *Second*, he suggests that these judgments display a "moral capacity" (or a "moral sense") just as, analogously, a child's ability to distinguish well-formed from incorrect sentences indicates a "grammatical sense."⁴ *Third*, Rawls observes that there are recognizably better and worse ways of making moral judgments. *Sound* methods of moral deliberation result in what Rawls terms "considered moral judgments." The first point is, I trust, sufficiently clear and obvious as to require little further elaboration. The other two points are of considerable importance, as we shall see.

The significance to Rawls of ordinary "moral sense" and the particular judgments that arise therefrom can scarcely be overrated. While this sense and these judgments do not supply final and conclusive criteria of justification for his theory, they do suggest useful points of departure, as Rawls indicates in the following passage:

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

Now one may think of moral philosophy 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 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. (p. 46)

Rawls's third point concerning ordinary moral deliberation is that not all judgments are of equal worth. Some have stronger claims upon us than others, and for reasons that we can readily recognize. For instance, we are more justified in accepting the judgment of a moral agent who is intelligent, well-informed, unbiased, and cool-headed, than by another who clearly lacks these traits. In other words, says Rawls, "considered moral judgments" are simply:

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. (p. 47-48)

Common, workaday moral insight and intelligence has a voice and a vote in justice as fairness.⁵ However, as we shall soon see, "considered moral judgments" do not, of themselves, have the power of decree or of the veto in Rawls's theory of justice.

"Reflective equilibrium" between judgments and principles. Very well, just where does all this leave us? Are the abstract rules of morality (the "principles") simple generalizations from particular, practical, moral decisions ("judgments"), or are moral judgments deductive applications of the general principles? And what of the contractual "original position?" Are the conditions of the original position to be interpreted, and perhaps adjusted, so that the contractors will be led to accept principles that "accommodate our firmest convictions?" Or, on the other hand, might not the resulting principles be sufficiently firm to require us to reconsider, and occasionally abandon, some "considered moral judgments?" In short, in justice as fairness, which has precedent: the principles derived from the conditions of the original position, or the considered moral judgments of practical life? Rawls, quite frankly, refuses to make a clear choice between these alternatives.

In searching for the most favored description of this situation [in the original position] we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions.⁶ We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. (p. 20)

All this, however, does not quite settle the matter, since there might be several competing modes of balancing general principles and practical judgments. On what grounds is one scheme of "reflective equilibrium" to be preferred to another? The criteria are not unlike those applied to scientific theories; namely, *simplicity*, *elegance*, *explanatory force*, and *coherence*. In no case should appeals to "intuition," "self-evidence," or "first principles" supplant evidence or systematic integrity. Rawls feels that Justice as Fairness meets these conditions:

It represents the attempt to accommodate within one scheme both reasonable philosophical conditions on principles as well as our considered judgments of justice. In arriving at the favored interpretation of the initial situation there is no point at which an appeal is made to self-evidence in the traditional sense either of general conceptions or particular convictions. I do not claim for the principles of justice proposed that they are necessary truths or derivable from such truths. A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view. (p. 21, see also pp. 121-122)

It is my primary purpose in this chapter to give a rather straightforward and sympathetic account of Rawls's theory, while directing little attention to his supporting arguments or the many critical responses to his book. However, a brief consideration of two recurring objections to "Justice as Fairness," and to "reflective equilibrium" in particular, will, I think, help to clarify Rawls's objectives and to suggest some of the new directions in moral philosophy indicated by his theory.

First objection: the theory is circular and uses ad hoc justifications. At this point the critical reader is likely to become suspicious of Rawls's scheme. After all, isn't this ready willingness both to change principles to suit judgments, and to revise judgments to suit principles plan and simple *circular reasoning*? And note also this remarkably candid admission by Rawls: "We want to define the original position so that we get the desired solution" (p. 141, see also p. 166). Isn't he revealing a ready willingness to make *ad hoc* adjustments in his theoretical model in order to accommodate his preferred conclusions? Considerations such as these have led R. M. Hare (1973) to the unflattering conclusion that Rawls's contractors:

. . . come to decisions that they come to simply because they are replicas of Rawls himself with what altruism he has removed and a veil of ignorance clapped over his head. It is not surprising, therefore, that they reach conclusions which he can accept. (July, p. 249)⁷

These are serious charges. How might Rawls respond? I would suggest that the best defense would include a qualified admission that his theory is, in fact, somewhat circular. The defense might follow with the bold proposal that circularity need not be *prima facie* grounds for dismissing the theory. It seems to me, in fact, that Rawls's theory draws a *large, inclusive, and open* circle around such diverse phenomena and concepts as "the moral sense," practical moral judgments, the theory of rational choice, a conceptual model (i.e., "the original position"), and a system of abstract moral principles. But most of these components of the "circle" are themselves open to revision and refinement in the

face of new facts, new experiences, conceptual clarification, and other data and procedures that might be brought in from "outside" the circle to advance the theory. New developments in some portions of the systemic "circle" (for instance, new information concerning moral psychology, or new conceptions of universal human needs), may lead to revisions and reformulations in other aspects of the theory (such as new conditions in the original position). It is hard to fault a theory that is designed to respond *in toto* to improvements in its parts. Indeed, this sort of "circularity" (if, indeed, we should properly call it that:), is characteristic of a vital and growing science.

The philosophers' long-standing aversion of circular argument is well-taken, if an offending argument draws a *closed* circle around a few identifiable components, and if these parts, like a pair of inebrates, hold each other up with no discernable outside support. On the other hand, an over-scrupulous and uncritical attention to traditional "rules of inquiry" may stultify inquiry. "Circularity" need not, in all cases, be sufficient grounds to reject an argument. (In §37, above, I will suggest that Rawls's attempt to avoid a similar philosopher's bugaboo, "the infinite regress fallacy," causes him some unnecessary anxiety.)

A final rejoinder to the charge of "circularity" might be: "what is the alternative?" If moral judgments are to be deduced entirely from general principles, without these principles being shaped, in turn, by "feedback" from moral experience, then the result will be moral *a priorism*, pure and simple; i.e., a type of "ethical Cartesianism." (Rawls categorically rejects this approach, see pp. 577-578.) If, conversely, the "principles" are nothing but generalizations drawn from a myriad of particular judgments, presumed to contain all moral wisdom, we are left with simple *subjectivism* and *intuitionism*, and thus, with no valid means of adjudicating moral disputes. Neither of these results is acceptable to Rawls (although, as we shall shortly discover, he has been charged with accepting the latter result) . Instead, he insists, each part of the moral theory must be capable of affecting, and being affected by, the other components. The parts, that is to say, must be in "reflective equilibrium" with each other.

Another objection: the theory is intuitive and subjective. Rawls believes that "reflective equilibrium" between principles, judgments, and the conditions of the original position, spare Justice as Fairness from the pitfalls of *intuitionism* and *subjectivism*. R. M. Hare (1973), however, is convinced that Rawls's rescue attempt has failed. Rawls, he insisted:

Is . . . advocating a kind of subjectivism, in the narrowest and most old-fashioned sense. He is making the answer to the question 'Am I right in what I say about moral questions?' depend on the answer to the question, 'Do you, the reader, and I agree in what we say?' This must be his view, if the considered judgments of author and reader are to occupy the place in his theory which is occupied in an empirical science by the facts of observation. Yet [on page 516] he claims objectively for his principles .

. . . The element of subjectivism enters only when a philosopher claims that he can 'check' his theory against his and other people's views, so that a disagreement between the theory and the views tells against the theory. To speak like this (as Rawls does constantly throughout the

book) is to make the truth of the theory depend on agreement with people's opinions. (April, pp. 145-146)

This, charges Hare, "is how phrases like 'reasonable and generally acceptable' are often used by philosophers in lieu of argument" (p. 145).

A related charge by Hare is that Rawls relies upon "scores" of intuitions ("almost always a form of disguised subjectivism"), many of which Hare cites explicitly. He continues:

Since the theoretical structure is tailored at every point to fit Rawls's intuitions, it is hardly surprising that its normative consequences fit them too – if they did not, he would alter the theory; . . . and the fact that Rawls is a fairly typical man of his times and society, and will therefore have many adherents, does not make this a good way of doing philosophy. (April, 1973, p. 147. Hare cites examples on pp. 19 and 141 in Rawls.)

In a recent paper, Spencer Carr (1975) replied, on Rawls's behalf, to Hare's serious attacks. *First*, Carr reiterates Rawls's persistent point that the allegedly "intuitive" and "subjective" considered moral judgments are *data*; the *beginning*, but not the *end*, of the theory. These beliefs, Carr reminds us, "are never treated as if they needed no support." In point of fact: "They require rational justification [and], . . . they can be overturned" (p. 93). *Second*, "our moral intuitions are an important part of our total moral experience, and a full moral theory cannot just ignore them. When they are not reliable, we will need to explain why this is so, drawing presumably upon both moral and psychological theory" (p. 93). *Third*, Hare's preferred approach to moral philosophy through the analysis of moral terms is not, itself, free from subjective factors. Writes Carr:

Insofar as one wants to exclude at the outset reference to moral intuitions, one is committed to the claim that we can derive substantial moral conclusions from the definitions and logic of moral terms, plus factual data. But one is also committed to the dubious claim that our willingness to accept a set of definitions and associated logic is not itself rooted in a set of moral intuitions. (p. 94)

Finally, we can ask once more, as we did in the case of the "circularity" objection, "*What is the alternative?*" – in this case, the alternative to employing the subjective data of ordinary moral deliberations? Is it *A Priorism*? Stipulations? Emotivism? Conventionalism? The first available moral autocrat? Indeed, if we do, in fact, restrict our moral investigation to analyses of "the logic of ordinary moral discourse," might we not be tacitly accepting conventional views embedded in ordinary language? Rawls is unwilling to take such risk as these. He feels, instead, that an ethical theory which scrupulously evaluates and systematizes elements of conceptual clarification, practical experience, the social sciences, the moral sense ("intuition?"), will be better equipped to answer our most basic and enduring moral questions.

There remains for Rawls's theory, one trap from which I can see no easy escape. On the one hand, Rawls offers a theory of justice, the principles of which are held to be in "reflective equilibrium" with

"considered moral judgments" of practical moral experience. However, as every freshman anthropology student must know, the judgments of practical experience are culture-dependent and, thus, vary widely from culture to culture. And yet, on the other hand, Rawls claims that Justice as Fairness has universal application. Can Rawls have it both ways? Parekh (1972) thinks not:

The test of reflective equilibrium is inconsistent with a "universalist" view of the theory of justice. This test ensures that what Rawls's theory does is explicate the sense of justice dominant among those whose judgments are taken into account, systematize the principles that they find natural and attractive. There is no more reason to believe that one theory of justice would satisfy all cultures than there is to believe that all cultures share the same judgments of justice. Rawls provides a theory which answers only to our judgments, a theory of justice for [Western democracies]. (p. 317)

The most graceful resolution of this dilemma, suggests Parekh, would be to abandon the claim of universality. He writes:

There is no scandal here. I do not see what is wrong with saying that the sense of justice by which we order – or at least criticize – our society reflects the image of man we have created and not an eternal human essence. (p. 323)

I would be most interested to read Rawls's reply to this criticism.⁸

23. The Original Position

We have arrived, at last, at the heart of Rawls's theory of justice: the original position -- the hypothetical assembly of rational egoists whose portentous task it is to decide, for all persons in all generations, the principles of justice that are to define right personal conduct and just institutional activity. Now it seems axiomatic that we can draw no conclusions whatever concerning the final decisions of this committee unless we understand, quite clearly, the conditions of their deliberations; i.e., the available knowledge, the general purpose of the assembly, the formal constraints upon their conclusions, and the motives of the parties on the committee. This section will present Rawls's conception of these operative conditions of the original position.

Before we proceed with the detailed account of the conditions of the original position, we surely have a right to ask, quite bluntly, what the point might be of this elaborate conceptual apparatus. Rawls perceives the original position as an "expository device" (p. 21), a thought-experiment designed (a) to include procedures of moral decision-making that are widely believed to be "fair," (b) to allow information generally believed to be relevant to such deliberations, (c) to exclude information that is morally irrelevant and which might provoke bias and special pleading.⁹ A conference so devised, says Rawls, would formulate and agree to "principles of justice . . . which rational persons concerned to

advance their interests would consent to as equals when none are known to be advantaged or disadvantaged by [morally arbitrary] social and natural contingencies" (p. 19).

For reasons which we will shortly encounter, the conditions of the original position cannot be fully duplicated in the real world; that is to say, no such "conference" could actually be convened, and thus the conclusions thereof are never actually and explicitly agreed to by identifiable persons. Why, then, should the results of this hypothetical assembly be of any interest to us? The reason, quite simply, is that:

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. Thus what we shall do is to collect together into one conception a number of conditions on principles that we are ready upon due consideration to recognize as reasonable. These constraints express what we are prepared to regard as limits on fair terms of social cooperation. One way to look at the idea of the original position, therefore, is to see it as an expository device which sums up the meaning of these conditions and helps us to extract their consequences. (p. 21. See also p. 587)

In addition to the question of the *content* of the conditions in the original position, there is the question of the "strength" or "weakness" of these conditions. This issue requires clarification. As noted earlier, *some* conditions in the original position must be specified; otherwise, there will be no resulting principles. On the other hand, if the stated conditions are too "strong," that is to say, too complex, detailed, and controversial, the scheme loses its ready plausibility and, thus, its claim to universality. Accordingly, Rawls seeks a "constrained minimum;" "a set of weak conditions that still enable us to construct a workable theory of justice" (p. 583, see also, note 6, p. 8, above).

The circumstances of justice. While rules of justice are applicable to most existing human associations, this is a contingent fact and not a logical truth. In a manna-from-heaven paradise or a community of stoics, wherein the supply of goods far exceeds demands, patterned schemes of cooperation and distribution are superfluous, as are rules for adjudicating competing claims. Similarly, a community of saints each united in their active and selfless devotion to common ideals would, in the absence of competing interests, have no need for rules of justice. Alas, the real world is rarely like this. On the other hand, conditions could be far worse than they are, in fact. Mankind has, for the most part, risen above the conditions of the Hobbesian state of nature: men are usually willing to act with a modicum of civility and to engage in joint enterprises for mutual benefit. Between these extremes of utopia and savagery -- which is to say, in virtually all known conditions of human society -- are to be found the "circumstances of justice," described by Rawls as "the normal conditions under which human cooperation is both possible and necessary" (p. 126).

While Rawls briefly notes several "circumstances of justice" (including such obvious conditions as physical proximity), he is concerned with two in particular: *moderate scarcity* and *mutual disinterest*. By "moderate scarcity," he means that:

. . . natural and other resources are not so abundant that schemes of cooperation become superfluous, nor are conditions so harsh that fruitful ventures must inevitably break down. While mutually advantageous arrangements are feasible, the benefits they yield fall short of the demands men put forward. (p. 127)

The parties in the original position must, therefore, understand that the principles that they devise are to apply to such circumstances. By "mutual disinterest" Rawls means that while individuals in a society may cooperate for mutual advantage their primary concerns are focused upon their own personal "life plans" pursued in behalf of their own private interests. Rawls does not deny that men are often motivated by genuine sentiments of altruism and affection. However, he does not wish to have his theory rest upon such occasional, however desirable, human traits. Accordingly, in the original position:

The postulate of mutual disinterest . . . is made to insure that the principles of justice do not depend upon strong assumptions. Recall that the original position is meant to incorporate widely shared and yet weak conditions. A conception of justice should not presuppose, then, extensive ties of natural sentiment. At the basis of the theory, one tries to assume as little as possible. (p. 129)

Rawls makes one crucial exception to the rule of "mutual disinterest." He stipulates that the parties in the original position might be thought of as "heads of families," and thus, concerned about the welfare of certain members of the generation which follows them. If this "heads of families condition" (as I shall call it) seems conspicuously arbitrary, I can only agree. In the following chapters, Rawls's position on this, and other conditions pertaining to posterity, will be carefully explicated and criticized.

The formal constraints of the concept of right. The simple requirement that the parties of the original position are to formulate *principles of justice* suggests immediate constraints upon the conditions of their deliberation and the range of acceptable alternatives. These constraints follow, not simply (as many contemporary philosophers would urge) from the concepts of *justice* or *right*, or from an analysis of the basic function of principles of right, namely, "adjusting the claims that persons make on their institutions and one another" (p. 131). Rawls's meta-ethical position is quite clear: "The merit of any definition depends upon the soundness of the theory that results; by itself, a definition cannot settle any fundamental question" (p. 130).¹⁰

Rawls identifies five constraints or criteria of the principles of right. *First*, such principles should be general: "that is, it must be possible to formulate them without the use of what would be *intuitively* recognized as proper names, or rigged definite descriptions" (p. 131). *Second*, the principles should be *universal*: "they must hold for everyone in virtue of their being moral persons" (p. 132). In a manner reminiscent of Kant's categorical imperative, Rawls elaborates: "a principle is ruled out if it would be self-contradictory, or self-defeating, for everyone to act upon it. Similarly, should a principle be reasonable to follow only when others conform to a different one, it is also inadmissible"

(p. 132). A *third* constraint requires that principles be public – that each morally responsible member of the community understand and acknowledge them, and furthermore understand that all others are similarly informed (p. 133). The *fourth* condition "is that a conception of right must impose an ordering on conflicting claims" (pp. 133-134). The *fifth* and last condition is *finality*:

The parties are to assess the system of principles as the final court of appeal in practical reasoning. There are no higher standards to which arguments in support of claims can be addressed; reasoning successfully from these principles is conclusive. (p. 135)

Thus, the principles of justice are to:

. . . override the demands of law and custom, and of social rules generally. We are to arrange and respect social institutions as the principles of right and justice direct. Conclusions from these principles also override considerations of prudence and self-interest. (p. 135)

To recapitulate: "a conception of right is a set of principles, *general* in form and *universal* in application, that is to be *publicly recognized* as a *final* court of appeal for *ordering* the conflicting claims of moral persons" (p. 135). (I have italicized the five "constraints.") These five conditions of right constrain the deliberation of the parties in the original position, as they formulate the principles of justice. As Rawls observes:

If the principles of justice are to play their role, that of assigning basic rights and duties and determining the division of advantages, these requirements are natural enough. Each of them is suitably weak and I assume that they are satisfied by the traditional conceptions of justice. These conditions do, however, exclude the various forms of egoism . . . which shows that they are not without moral force. This makes it all the more necessary that the conditions not be justified by definition or the analysis of concepts, but only by the reasonableness of the theory of which they are a part. (p. 131)

We have determined that the parties in the original position are to understand that they are to live under the basic "circumstances of justice"; that is to say, they will be "mutually disinterested persons" who will "put forward conflicting claims to the division of social advantages under conditions of moderate scarcity" (p. 128). The condition of "mutual disinterest" applies as well within the original position itself. In addition, the parties understand that the principles that they adopt must meet the five formal criteria of the concept of right that I have just enumerated. However, we have not yet fully discussed the question of what information may or may not be allowed in the deliberations by the parties, nor have we examined the issue of their motivations. We will turn next to these conditions.

The veil of ignorance. "Justice" is often symbolically personified as a blindfolded woman holding a balance scale. The scale, of course, signifies equality before the law -- a concept which, as we have seen, Rawls fully endorses. The blindfold indicates that the process of justice should be *unbiased*; that is, uninfluenced by irrelevant information or by special personal interest and circumstances. In courts of law, information brought before the jury is restricted by rules of admissible evidence and by such

procedures as objections by attorneys, sequestering of the jury, and so on. Rawls is similarly concerned that the parties of the original position not be swayed in their deliberations by irrelevant or self-serving considerations. Accordingly, he insists that the principles of justice be chosen "behind a veil of ignorance." Thus, in the original position:

. . . no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities. (p. 12)

In addition, the parties have no knowledge of the course of history (p. 200), and "no information as to which generation they belong" (p. 137), a condition which will prove to be of great significance, later in this dissertation. In general, we can say that the parties in the original position have no knowledge whatever of their distinguishing personal qualities, preferences, or circumstances in "real life." They are, however, aware of a few conditions that will affect them in some unknown age, culture and circumstance, once the veil is lifted.

Why this strange and, practically speaking, unattainable set of conditions? Rawls replies:

The idea here is simply to make vivid to ourselves the restrictions that it seems reasonable to impose on arguments for principles of justice, and therefore on those principles themselves. Thus it seems reasonable and generally acceptable that no one should be advantaged or disadvantaged by natural fortune or social circumstances in the choice of principles. It also seems widely agreed that it should be impossible to tailor principles to the circumstances of one's own case. We should insure further that particular inclinations and aspirations, and person's conceptions of their good do not affect the principles adopted. The aim is to rule out those principles that it would be rational to propose for acceptance, however little the chance of success, only if one knew certain things that are irrelevant from the standpoint of justice. (p. 18)

The exclusion of knowledge of one's personal "conception of the good," that is, of personal aims, preferences, and life plans, may seem to be particularly puzzling. The point of this condition should be evident if we recall that, as a deontologist, Rawls perceives the *right* to be prior to the *good*. It follows that Rawls will require that the "just" rules of association, chosen in the original position, "put limits on which satisfactions have value" and that, in effect, the parties agree "to conform their [various personal] conceptions of their good to what the principles of justice require" (p. 31). To assure this priority of the *right*, conceptions of personal good must be excluded from those very deliberations which are to decide principles of right conduct. But can *any* decision be made in the original position without *some* conception of what is required, and thus desirable, for a tolerable life? Rawls acknowledges that some minimal conception is necessary, and thus allows the parties to understand what it is that "a rational man wants, whatever else he wants" (P. 92). This conception of basic, necessary "primary goods," which supplies the motivational factor in the original position, will be discussed shortly.

Admissible knowledge. What, then, are the parties in the original position permitted to know? Concerning themselves (that is to say, concerning *all* members of the assembly *equally* and *collectively*), they are advised: (a) that they represent continuing family lines, and (b) that they all belong to the same generation in actual life, whatever generation this may in fact be (the "present time of entry condition"). These two conditions, of course, have important bearing upon the question of the duty to posterity, and thus will receive close attention later in this dissertation. Returning to the parties' knowledge of their circumstances beyond the veil of ignorance: (c) they understand, as we have noted, that their society is subject to the "circumstance of justice" (i.e., moderate scarcity and mutual indifference), (d) they are aware of, and wish to secure for themselves, certain "primary goods" -- things that they will need and desire, whatever else they may find that they want once the veil is lifted. Finally, and this is a crucial condition (e), they understand that they must be prepared to live according to the principles of justice that they choose, whatever may be the condition of their life, their society, or their generation, beyond the veil of ignorance. That is about all that they know about themselves, and about each other. Beyond this, however, there are virtually no restrictions upon general knowledge:

They know the general facts about human society. They understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology. Indeed, the parties are presumed to know whatever general facts affect the choice of the principles of justice. There are no limitations on general information, that is, on general laws and theories, since conceptions of justice must be adjusted to the characteristics of the systems of social cooperation which they are to regulate, and there is no reason to rule out these facts. (pp. 137-138)

Thus, we are reminded, once again, that Justice as Fairness is a fact-contingent theory of justice (Cf. §20, above).

Primary goods. In the original position, no one is permitted to know the particular circumstances of his actual life -- not even his own conception of the good life, which is to say the "life plan," that he has devised to achieve his preferred ends and interests. And yet, it is the task of the parties in the original position to choose principles of justice that are to "[assign] basic rights and duties and [determine] the division of advantages" (p. 131). How can this task be accomplished, unless it can be determined which of the alternative conceptions of justice will yield the most advantages?; and how can "advantages" be weighed without some notion of *ends*, which is to say of *goods*, which the parties in the original position will be motivated to secure for themselves? Rawls acknowledges that this challenge cannot be dismissed. He thus concedes that, if the principles of justice are to be established, "it is necessary to rely on some notion of goodness, for we need assumptions about the parties motives in the original position" (p. 396). However, realizing that he is thus jeopardizing his deontological position (that is to say, the primacy of the *right*), he adds: "the theory of the good used in arguing for the principles of justice is restricted to the bare essentials" (p. 396). This theory of "primary goods," that is, of goods desired by any rational person, whatever else he might want, is called, by Rawls, "the thin theory of the good."

Very well, just what *are* the primary goods? Obviously, a list of such goods should include such "natural" desiderata as health, vigor, intelligence and imagination. However, these goods are of less interest to Rawls since, he believes, their attainment is not normally the direct result of the rules of justice that obtain in a society (p. 62). Of foremost concern to Rawls, and therefore to us, are the "social primary goods," the distribution of which he believes are contingent upon the principles of justice adapted in the original position. These goods include self-respect, "rights and liberties, powers and opportunities, income and wealth" (p. 62). And from whence does Rawls acquire the list of primary goods? It can, he says:

. . . be accounted for by the conception of goodness as rationality in conjunction with the general facts about human wants and abilities, their characteristic phases and requirements of nurture, . . . and the necessities of social interdependence. (p. 434)

As this exposition proceeds, we will have many occasions to cite these "primary goods," separately, and in combination.

It may seem significant that *benevolence* is *not* listed as a primary good, a circumstance of justice, or any other type of motivating factor in the original position. While Rawls does not dispute the manifest advantages to society of widespread sentiments of good will among its members, he chooses not to have the stability of the principles of justice depend upon this sentiment. This choice is consistent with his preference for "weak" conditions in the original position. Instead, Rawls suggests that the combination of mutual disinterest and the veil of ignorance will have the same motivational result as benevolence. In other words, since the parties do not know who they will be in actual life, they will adopt rules of generosity, forbearance, and mutual aid toward others in the knowledge that, in a sense, they may very well turn out to *be* these "others" (pp. 148-149). Peter Caws (1972) vividly illustrates this point:

If one actually ran the risk of *being* a heroin-addicted veteran, a woman with an unwanted pregnancy, a black dropout, one's attitude to drugs or abortion or ghetto schools might be different. So a rule-of-thumb test for the justice (and indeed for the morality) of an interpersonal action becomes: Would I act on the same principle if I did not know what my own position relative to that of the other person involved? (p. 24)

This amounts, in effect, to an approximate restatement of "the golden rule."

In the original position, then, the rational strategy is to maximize one's own prospects in actual life:

The persons in the original position try to acknowledge principles [of justice] which advance their system of ends as far as possible. They do this by attempting to win for themselves the highest index of primary social goods, since this enables them to promote their conception of the good most effectively whatever it turns out to be. The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor. Nor

do they try to gain relative to each other; they are not envious or vain. Put in terms of a game, we might say: they strive for as high an absolute score as possible. (p. 144)

And since they may turn out to be anyone, it is rational for them to enhance the prospects and minimize the risks *for everyone*. (The recurrence of this phrase "maximizing one's prospects," and its cognates, may suggest that Justice as Fairness is a variant of Utilitarianism. However, as I will point out in §27, there are fundamental and irreconcilable differences between the two theories.)

Rawls believes that it is possible to define the "primary goods" with sufficient generality that they will apply to all persons in all ages, cultures, and circumstances. With this universal "thin theory of the good," says Rawls, Justice as Fairness:

Sets up an Archimedean point for assessing the social system without invoking *a priori* considerations. The long range aim of society is settled in its main lines irrespective of the particular desires and needs of its present members. And an ideal conception of justice is defined since institutions are to foster the virtue of justice and to discourage desires and aspirations incompatible with it. (p. 261)

Does Rawls succeed? For that matter, is such an accomplishment attainable in principle? Several recent papers have been addressed to these questions.¹¹ Indeed, Rawls's claim to time and culture neutrality may be one of the most controversial aspects of his theory. (See also note 8, p. 92, above).

The rules of deliberation. The parties in the original position understand, then, that they are choosing principles that will regulate the conditions of their association, under conditions of moderate scarcity and mutual disinterest. According to the constraints of the concept of right, they know that these principles of justice must be general, universal, public, ordered, and final and binding. They have at their disposal all general knowledge that may be relevant to their deliberations. However, with a few exceptions previously noted (e.g., the "time of entry" and "heads of families" conditions), they have no particular knowledge of themselves. They do, however, know that they will want to choose a set of principles that will maximize their allotments of "primary goods" and minimize their risks, once the veil of ignorance is lifted and they find themselves in the conditions of their actual lives.

Under such circumstances, how will the parties deliberate? *First*, they agree that they must rely on each other to adhere to the principles chosen. Accordingly, the principles must be stable and acceptable; that is to say, they must survive the "strains of commitment." Therefore, the parties "cannot enter into agreements that may have consequences that they cannot accept" (p. 176). *Second*, since the decisions are final for all time and are to "govern [their] life prospects," the parties must not only "protect their basic rights," they must also "insure themselves against the worst eventualities" (p. 176). *Third*, the grave import of these decisions requires no less than *unanimity*. But, of course, with no knowledge of distinguishing personal characteristics and circumstances, the parties are as alike as coins from the same mint or stamps from the same sheet. Unanimity is thus assured. *Fourth*, it also follows that "bargaining" for advantages will be useless, since "no one knows his situation in society nor his natural assets, and therefore no one is in a position to tailor principles to his

advantage" (p. 139). And what is to be the *content* of their decisions? I will explain Rawls's answer to this crucial question in the next section.

24. The Principles of Justice

The original position, that marvelous, complex, and imaginative conceptual apparatus, has been assembled by Rawls to fulfill a simple and fundamental function: to bring forth well-articulated, well-ordered, and well-argued *principles of justice*. These are principles "which rational persons concerned to advance their interests would accept in this position of equality to settle the basic terms of their association" (pp. 118-119). These "terms" are likewise to apply to the social, political, and economic institutions of their society. Conversely, it follows that the resulting principles of justice may be interpreted as "the solution for the problem of choice presented by the original position." In other words, "given the circumstances of the parties, and their knowledge, beliefs, and interests, an agreement on these principles is the best way for each person to secure his ends in view of the alternatives available" (p. 119). These principles are to apply to two basic aspects of the social order: (a) "they are to govern the assignment of rights and duties," and (b) they are "to regulate the distribution of social and economic advantages" (p. 61).

Through a series of subtle, complex, and elegant arguments, Rawls develops what he believes would be the principles of justice chosen in the original position as he devised it. It is unfortunate that space permits little more than a report of the results, with merely an occasional sketch of one or another supporting argument. Scores of papers have been written analyzing and criticizing details of Rawls's defenses of the principles; they must, for the most part, be ignored here. However, in the following chapter, I will offer a more detailed account of Rawls's position concerning the particular issue of "justice between generations" and of his defense thereof. The remaining chapters of the dissertation will be devoted to a critical analysis of Rawls's position concerning posterity.

"Representative persons" and the principles of justice. I have noted that the parties of the original position seek to devise rules of justice that will maximize their expectations. However, before we proceed with an explication of these principles, a brief qualification is in order. Each member of the original position, we will recall, has no knowledge of his distinctive personal traits and circumstances. He is not, that is to say, an identifiable person. Furthermore, the principles chosen must meet the formal requirement of generality; i.e., they must not be variously applicable to specific persons. It follows, therefore, that the principles of justice must not apply "to distributions of particular goods to particular individuals who may be identified by their proper names" (p. 64). Accordingly, "when principles mention persons," it must be understood that "the reference is to representative persons holding the various social positions, or offices, or whatever, established by the basic structure" (p. 64). (Note Rawls's vagueness here!)

In addition, Rawls believes that distributive arrangements in a society are generally "close knit"; that is to say, "the expectations of representative persons depend upon the distribution of rights and duties throughout the basic structure." Consequently, "expectations are connected: by raising the prospects

of the representative man in one position we presumably increase or decrease the prospects of representative men in other positions" (p. 64). As will soon be apparent, the interconnectedness of expectations among representative persons bears important implications for rules of distributive justice.

The principle of equal liberty. Rawls contends that *liberty* is a basic prerequisite to the attainment and maintenance of the social primary goods. He gives the following description of "liberty": "This or that person (or persons) is free (or not free) from this or that constraint (or set of constraints) to do (or not to do) so and so" (p. 202). Accordingly, "persons are at liberty to do something when they are free from certain constraints either to do it or not to do it and when their doing it or not doing it is protected from interference by other persons" (p. 202). Liberty, or its restriction, applies to associations as well as individuals, and "constraints may range from duties and prohibitions defined by law to the coercive influences arising from public opinion and social pressure" (p. 202). (Cf. my discussion of "rights," §7, above.) Rawls's primary interest is the connection of liberty to "constitutional and legal restrictions." "In these cases," he writes, "liberty is a certain structure of institutions, a certain system of public rules defining rights and duties" (p. 202).

Equal liberty is necessary in a society based upon the principles of justice in that it assures that all citizens will have, and acknowledge, a personal stake in the maintenance of the just order. Liberty also assures that each member of the community will have the opportunity to pursue his personal plan of life without undue interference. Because the parties in the original position do not know what tastes, goals, and ideals they will have beyond the veil, they will insist upon an equal liberty of conscience, since "they cannot take chances with their liberty by permitting the dominant religious or moral doctrine to persecute or to suppress others if it wishes" (p. 207). Their life-plans, whatever they turn out to be, must not be stifled. Still less are the parties willing, as representatives of continuing generations, to bargain away the free conscience of their descendants (p. 206).

Equal liberty brings even more immediate dividends. Among the most prominent is the enhancement of self-respect, one of the most significant of the primary goods. And why should "self-respect" be so important to Rawls? His answer is both clear and emphatic:

First, [self-respect or self-esteem] includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect implies a confidence in one's ability, so far as it is within one's power, to fulfill one's intentions. When we feel that our plans are of little value, we cannot pursue them with pleasure or take delight in their execution. Nor plagued by failure and self-doubt can we continue our endeavors. It is clear then why self-respect is a primary good. Without it nothing may seem worth doing, or if some things have value for us, we lack the will to strive for them. All desire and activity becomes empty and vain, and we sink into apathy and cynicism. Therefore, the parties in the original position would wish to avoid at almost any cost the social conditions that undermine self-respect. (p. 440)

Arbitrary, accidental, morally irrelevant, and (most significantly) morally unacceptable inequalities are prominent among the "social conditions that undermine self-respect." Conversely, a constant awareness and public acknowledgment that "each counts for one, and no one counts for more than one," extends and secures the primary good of self-esteem throughout the community (p. 545). Thus, says Rawls:

. . . the basis for self-esteem in a just society is . . . the publicly affirmed distribution of fundamental rights and liberties. And this distribution being equal, everyone has a similar and secure status when they meet to conduct the common affairs of the wider society. (p. 544)

In sum, an assurance of equal liberty "strengthen(s) men's sense of their own worth, enlarge(s) their intellectual and moral sensibilities, and lay(s) the basis for a sense of duty and obligation upon which the stability of just institutions depends" (p. 234).

Rawls concludes that the advantages of equal liberty are so great, and the dangers of its abridgment so grave, that the parties in the original position would mandate, above all others, the following principle of justice:

Each 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. 250)

The principle of equal opportunity. The principles of justice, we will recall, are to have two general roles: *first*, that of "assigning basic rights and duties," and, *second*, that of "determining the division of advantages" (p. 131). As noted, *rights* and corresponding *duties* to protect and sustain these rights, are to be assigned *equally*. However, the same need not be said for "the division of advantages." If an uneven distribution to some results in greater advantages to *all*, such an inequality of distribution is acceptable. Indeed, some (though not all) inequalities of distribution are *inevitable*. For instance, natural talents are unevenly distributed, as are certain social advantages (for example, the benefits of being nurtured in a stable and loving family). Nothing can be done to "even out" these unequal distributions; they are simple, unyielding, natural facts, and thus the rules of justice must be drawn to accommodate them. The rational strategy in the original position, therefore, would be to treat these uneven distributions as community assets:

. . . and to share in the benefits of this distribution whatever it turns out to be. Those who have been favored by nature, whoever they are, may gain from their good fortune only on terms that improve the situation of those who have lost out. (p. 101)

Since "no one deserves his greater natural capacity nor merits a more favorable starting place in society" (p. 102), it seems to follow that unequal distributions can *only* be accepted on grounds of their community benefit. Accordingly, those positions in society which bring greater advantages must be equally open to all applicants. In other words, "equal life prospects in all sections of society" are owed to all who are "similarly endowed or motivated" (p. 301). To insure fair and equal opportunity, then,

advantages should be distributed according to rules of "pure procedural justice" -- rules that are objective, public and blind to morally irrelevant personal traits.¹²

In summary: the parties in the original position, not knowing their social status or circumstances, or their drawing in the "lottery" of natural talents or family conditions, would choose a rule of equal opportunity to gain advantages. Furthermore, "fairness" requires that any unequal distributions of advantages be treated as community resources. Accordingly, the members of the original position would choose the following principle:

Social and economic inequalities are to be arranged so that they are . . . attached to offices and positions open to all under conditions of fair equality of opportunity. (p. 302)

The difference principle. The foregoing presentation of the equal opportunity principle has brought us well along toward an explanation of Rawls's third principle of justice: "The Difference Principle." (Strictly speaking, "equal opportunity" and "the difference principle" are sub-parts of Rawls's Second Principle of Justice.) In particular, we have noted that the advantages of native talents and family background are purely random, not deserved, and morally neutral. Accordingly, Rawls argues, these advantages should be treated as community assets. There are additional and familiar reasons why unequal distributions of advantages might be to the common good. For instance, some socially beneficial professions which require long and disciplined training (e.g., medicine), or extra hazards (e.g., law enforcement) may demand higher than average incentives or compensation if the community's need for these services is to be met.

Knowing these general facts, ignorant of their eventual circumstances in actual life, and yet seeking to secure for themselves the highest expectation of social primary goods, what rules will the parties in the original position choose to regulate the distribution of these goods? Rawls argues that they will choose the most conservative strategy: the "maximin" strategy which would allow the *maximum* possible expectations for the least favored (*minimum*) members of the community. In Rawls's words:

The . . . principles are those a person would choose for the design of a society in which his enemy is to assign him his place. The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcomes of the others. (pp. 152-153)

Why this strange and extreme distribution policy? *First*, Rawls contends, in the absence of any particular information concerning outcomes, it is prudent to discount any probability calculations. *Second*, achieving the minimum index of primary goods is vastly more important than any marginal benefits that may accrue from surplus goods. Thus, it is far better to avoid falling below the line than to risk such deprivation by gambling for higher stakes. *Third*, the parties agree that the worst possible outcomes (e.g., slavery or despotism) of other distributive schemes are not acceptable (p. 154). *Finally*, and most significantly for our particular issue, the "choice of principles should seem reasonable to others, in particular their descendants, whose rights will be deeply affected by it" (p. 155). After all, Rawls argues:

[We should be] more reluctant to take great risks for them than for ourselves; and we [should be] willing to do so only when there is no way to avoid these uncertainties, or when the probable gains, as estimated by objective information, are so large that it would appear to them irresponsible to have refused the chance offered even though accepting it should actually turn out badly. (p. 169)

The question of the distribution of goods is not confined to the rights and expectations of one's contemporaries. Resources and opportunities must also be set aside for succeeding generations. Thus, the least favored of one's own generation *and* the least favored of succeeding generations are, in a sense, competing for the minimum shares. This issue of "just savings" (as Rawls calls it) will be of pre-eminent concern to us in the remaining chapters of this dissertation. For the moment, I wish only to raise the issue and identify its place in Rawls's theory of justice.

How, then, would the maximin strategy be applied in the original position to the question of just distributions of primary goods? Rawls believes that the parties would choose the following "Difference Principle":

Social and economic inequalities are to be arranged so that they are . . . to the greatest benefit of the least advantaged, consistent with the just savings principle. (p. 302)

We have considered the argument for the difference principle from the point of view of the least advantaged member of society. According to this principle, any attempt to allot him more would result in his getting less. But what of the person who is better favored? After all, he would have to make do with less if his having more would be to the disadvantage of the least favored. Hasn't he a valid complaint? Rawls thinks not:

To begin with, it is clear that the well-being of each depends on a scheme of social cooperation without which no one could have a satisfactory life. Secondly, we can ask for the willing cooperation of everyone only if the terms of the scheme are reasonable. The difference principle, then, seems to be a fair basis on which those better endowed, or more fortunate in their social circumstances, could expect others to collaborate with them when some workable arrangement is a necessary condition of the good of all. (p. 103)¹³

Priority rules. It is not enough to list separate rules of justice. In practical affairs, principles come into conflict and, unless rules of priority are established, such disputes might not be arguable on any grounds more substantial than ad hoc appeals to "intuition," emotion, or some other subjective factors. Rawls does, in fact, rank his principles, with first priority to equal liberty, second priority to equal opportunity, and lowest priority to the difference principle.

This is helpful as far as it goes; but it scarcely goes far enough. There is little guidance here for setting cost-benefit balances. For instance, how little equal liberty gained will compensate for how much equal opportunity lost? How little opportunity gained is worth how much loss for the least advantaged? Rawls's answers to these questions are extreme: no amount of gain in social or economic

advantages (however great) can justly compensate for *any* amount of loss in equal liberty (however slight). Similarly, equal opportunity holds absolute priority over the difference principle. The principles are ranked, in Rawls's terms, "serially" or "lexically" (pp. 43, 61, 302-303).

The phrase "lexical ordering" requires some clarification. Suppose you desired to look up the name "Rawls" in *Who's Who* or the *Directory of American Philosophers*. You would first, of course, look under the "r's." Once you had located the "r's," you would look in the second letter-place for the "a's," again regardless of the succeeding letter in the name, and so on until you found the listing. In general, says Rawls, "lexical ordering":

Requires us to satisfy the first principle in the ordering before we can move on to the second, the second before we consider the third, and so on. A principle does not come into play until those previous to it are either fully met or do not apply. A serial ordering avoids, then, having to balance principles at all; those earlier in the ordering have an absolute weight, so to speak, with respect to later ones, and hold without exception. (p. 43)

Why such an extreme priority rule for ranking the principles of justice? First, the very fact that the priority of equal liberty is absolute affords each citizen a generous allotment of the primary good of self-respect. "Respect for persons," says Rawls:

. . . is manifest in the content of the principles to which we appeal. Thus to respect persons is to recognize that they possess an inviolability founded on justice that even the welfare of society as a whole cannot override. It is to affirm that the loss of freedom for some is not made right by a greater welfare enjoyed by others. The lexical priorities of justice represent the value of persons that Kant says is beyond all price. (p. 586)

Clearly, the respect for others, generated by the priority of liberty, is at least equaled by the respect for oneself that is likewise engendered thereby.

A second reason to accept a lexical ordering of the principles is that the right to pursue one's own plan of life -- to be, that is, the creator, evaluator, and executor of one's own destiny -- is of such paramount importance that without it any increased material advantages would be empty and without savor (pp. 542-543). Furthermore, the priority of liberty instills in each citizen a sense of loyalty to the institutions that embody this principle and, with this loyalty, a desire to protect and perpetuate these just institutions.

Even so, does all this justify *absolute* priority of liberty? Brian Barry (1973), among many others, feels that this rule is "outlandishly extreme" in that:

. . . as between two situations, the smallest superiority on the first principle outweighs any amount of superiority on the second principle and that the smallest amount of improvement on the first principle is worth sacrificing any amount of loss on the second principle. (pp. 59-60)

Rawls seems, in fact, to sense that he may have overreached a bit with his "lexical ordering" of the principles. Early in the book he writes: "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." (p. 45) What is *not* "clear" in the book is how we might unequivocally identify these "certain special though significant conditions."

Another hedge in the rule of "the lexical priority of liberty" is that the serial ranking of the principles applies only in conditions of *moderate* scarcity. In conditions of *acute* scarcity, it might well be more prudent (as Bertrand Russell once put it) to prefer a bushel of grain to the right to vote¹⁴ However, Rawls argues, this is a passing phase in history, for "as the conditions of civilization improve, the marginal significance for our good of further economic and social advantages diminishes relative to the interests of liberty" (p. 542)

Rawls, then, makes a fundamental distinction between the *general conception* of justice, which holds in cases of acute scarcity, and the *special conception* of justice (described in this section) which holds in conditions of, at worst, moderate scarcity. This distinction is based upon the assumption that:

It is only when social conditions do not allow the effective establishment of [basic liberties] that one can acknowledge their restriction. The denial of equal liberty can be accepted only if it is necessary to enhance the quality of civilization so that in due course the equal freedoms can be enjoyed by all. The lexical ordering of the two principles is the long-run tendency of the general conception of justice consistently pursued under reasonably favorable conditions. Eventually there comes a time in the history of a well-ordered society beyond which the special form of the two principles takes over and holds from then on. (p. 542)

At this stage in human progress, the principle of liberty properly assumes a lexical priority over the distributive principles, for:

As the conditions of civilization improve, the marginal significance for our good of further economic and social advantages diminishes relative to the interests of liberty, which become stronger as the conditions for the exercise of the equal freedoms are more fully realized. Beyond some point it becomes and then remains 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)

A summary statement of the principles of justice. This completes our all too brief account of Rawls's derivation of the principles of justice as they would be chosen from the standpoint of the original position. There remains only a complete statement in Rawls's words of these principles and their priority rules.

The General Conception of Justice -- (Condition of Acute Scarcity):

All social primary goods -- liberty and opportunity, income and wealth, and the bases of self-respect -- are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored. (p. 303)

The Special Conception of Justice -- (Condition of Moderate Scarcity):

First Principle: Each 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.

Second Principle: Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with the just savings principle, and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.

First Priority Rule (The Priority of Liberty):

The principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty.

There are two cases: (a) a less extensive liberty must strengthen the total system of liberty shared by, all; (b) a less than equal liberty must be acceptable to those with the lesser liberty.

Second Priority Rule (The Priority of Justice over Efficiency and Welfare):

The second principle of justice is lexically prior to the principle of efficiency¹⁵ and to that of maximizing the sum of advantages [i.e., classical utilitarianism]: and fair opportunity is prior to the difference principle.

There are two cases: (a) an inequality of opportunity must enhance the opportunities of those with the lesser opportunity; (b) an excessive rate of saving must on balance mitigate the burden of those bearing this hardship. (pp. 302-303)

The elements of Rawls's theory of justice: a schematic summary. Before moving on to a consideration of Rawls's views concerning institutional justice and the pursuit of personal good, I would like to draw a schematic representation of the relationships among the parts of Rawls's theory that we have examined so far. *First*, we will recall that, according to Rawls, right policies and judgments should be in accordance with the principles of justice, and that these principles, in turn, should follow from the deliberations of the parties in the original position as these deliberations are determined by the conditions and the admissible knowledge in that assembly. *Second*, Rawls believes that the considered moral judgments which are derived from rational, intelligent reflections upon experience and "the moral sense," should both affect and be affected by the formal theory of justice (described in the first point, above). *Finally*, the formal principles of justice and considered moral

judgments are brought into accord through "reflective equilibrium," whereby (a) the "considered judgment" might require alterations in the conditions of the original position, or (b) the resulting principles of justice, or entailed policies, might suggest revisions in the considered judgments. Thus, we arrive at the schematic diagram. Note that the uni-directional arrow from "considered moral judgment" to "conditions on the original position" does not mean that there is no influence in the opposite direction. However, the conditions in the original position may only affect the considered moral judgments *through* conflicts of these judgments with the principles of justice and/or the practical policies resulting from the conditions in the original position.

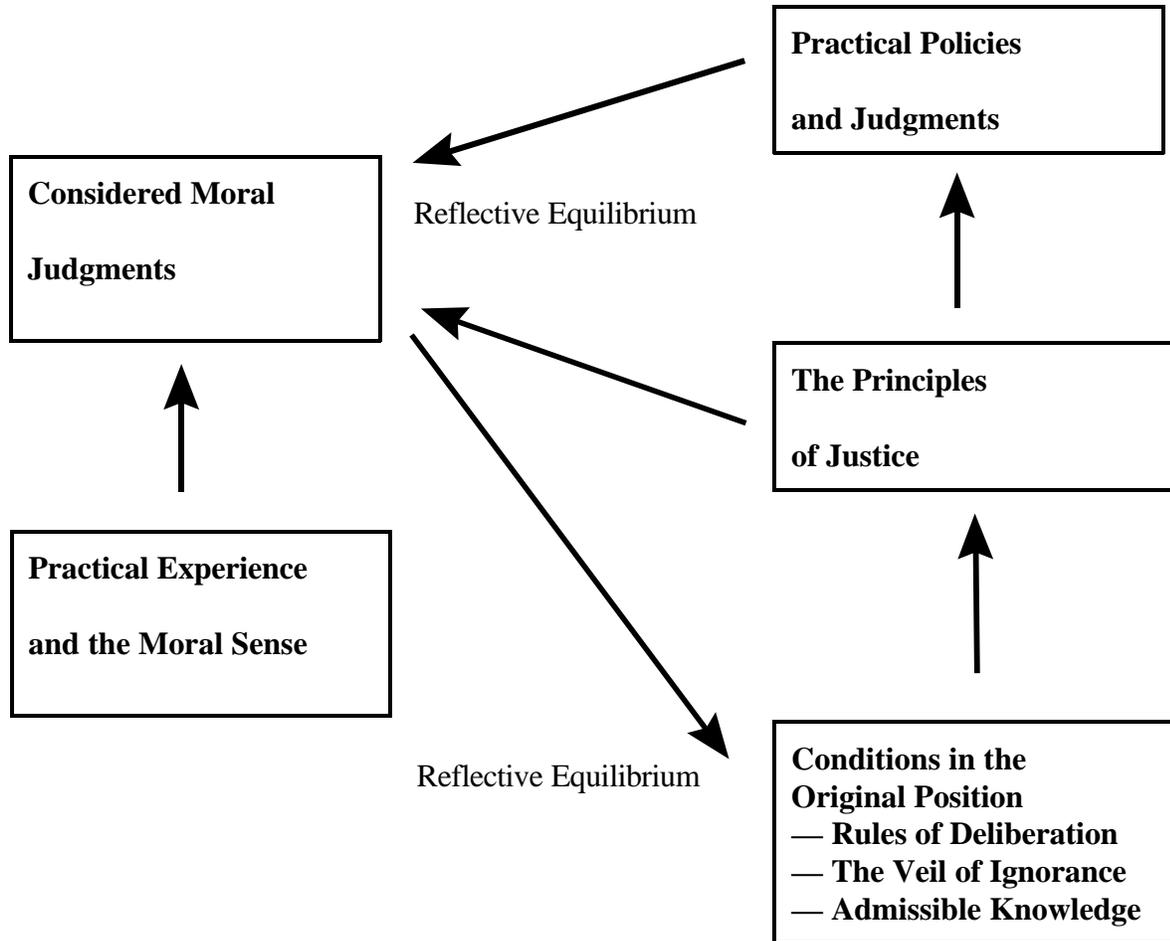


Figure 2. The Structure of "Justice as Fairness"

25. Just Institutions

Earlier (§20), I noted that Rawls considers "the basic structure of society" to be "the primary subject of justice" (Rawls, p. 7). I also pointed out that Rawls perceives the basic task of the parties in the original position to be the determination of "the fundamental terms of their association" (Rawls, p. 11). It is clear in this and numerous other passages in the book that Rawls wishes to articulate rules for the just function of *institutions*. Indeed, the second of the three main parts of the book is titled, "Institutions." It is noteworthy that Rawls's treatment of "Justice Between Generations" (§44) is to be found midway through this analysis of "just institutions," in his chapter on "Distributive Shares" (Chapter V).

Rawls interprets "institutions" to mean "a public system of rules which defines offices and positions with their rights and duties, powers and immunities, and the like" (p. 55). These rules, he continues:

Specify certain forms of action as permissible, others as forbidden; and they provide for certain penalties and defenses, and so on, when violations occur. As examples of institutions, or more generally social practices, we may think of games and rituals, trials and parliaments, markets and systems of property. An institution may be thought of in two ways: first as an abstract object, that is, as a possible form of conduct expressed by a system of rules; and second, as the realization in the thought and conduct of certain persons at a certain time and place of the actions specified by these rules. (p. 55)

The well ordered society. By describing an institution as "a public system of rules," Rawls reiterates a criterion that is found among his "formal constraints on justice"; namely, the publicity condition. Accordingly, a person engaged in just institutional activity:

. . . knows what he would know if these rules and his participation in the activity they define were the result of an agreement. A person taking part in an institution knows what the rules demand of him and of the others. He also knows that the others know this and that they know that he knows this, and so on. (p. 56)

If the institutional rules that are publicly acknowledged and obeyed include the principles of justice, the happy result may be called a "well ordered society." This important concept deserves some amplification.

In his recent paper, "Fairness to Goodness," Rawls (October 1975, pp. 547-548) acknowledges that in *A Theory of Justice* the basic features of the "well ordered society" "were not stated together at any one place" (p. 547n). He then proceeds to remedy the omission with a seven-item criterion list. Because the concept will figure significantly in later sections of this dissertation, I should like to present Rawls's list. After briefly characterizing the "well ordered society" as "one that is effectively regulated by a public conception of justice," Rawls continues that in such a society:

(1) Everyone accepts, and knows that others accept, the same principles (the same conception) of justice. (2) Basic social institutions and their arrangement into one scheme (the basic structure of society) satisfy and are with reason believed by everyone to satisfy these principles. (3) The public conception of justice is founded on reasonable beliefs that have been established by generally accepted methods of inquiry.

Furthermore, the members of a well ordered society, being free and equal moral persons:

(4) . . . have, and view themselves as having, a sense of justice (the content of which is defined by the principles of the public conception) that is normally effective (the desire to act on this conception determines their conduct for the most part). (5) . . . have, and view themselves as having, fundamental aims and interests (a conception of their good) in the name of which it is legitimate to make claims on one another in the design of their institutions. (6) . . . have, and view themselves as having, a right to equal respect and consideration in determining the principles by which the basic structure of their society is to be regulated.

Finally, the well ordered society is *stable*, which means that "(7) Basic institutions generate an effective and supporting sense of justice" (p. 547-548). (Cf. p. 112, below. See also Rawls, 1971, pp. 453-454.)

Political institutions and "the four-stage sequence. The principles of justice, which are to govern and set limits upon institutional activity in the well-ordered society, are formulated from the standpoint of the original position. However, this ideal and abstract perspective, by its very design, offers limited guidance for meeting specific problems of an actual society, since the practical work of operating institutions calls for particular knowledge that is excluded from the original position. On the other hand, the original position reflects the requirement that just decisions within institutional contexts should be arrived at judiciously; that is, through objective, impersonal, and rational decision-making procedures. How, then, can rational disinterest effectively be combined with the particular knowledge required for effective institutional functioning? Rawls's answer to this dilemma, as it applies to the governing political institutions of the society, is ingenious. He suggests that political institutions might be perceived as a series of assemblies which follow the adjournment of the original position, and at which the veil of ignorance is lifted in stages as questions of progressive specificity are dealt with (Cf. Rawls, §31).

The task of 'the original position, of course, is to choose the fundamental rules of association in *any* human community existing under the circumstances of justice. Since we have examined this conception in detail, I'll say little more about it. Rawls proposes that, upon deciding upon the principles of justice, the parties should immediately move on to a *constitutional convention*, at which time they will learn of the "relevant general facts about their society, that is, its natural circumstances and resources, its level of economic advance and political culture, and so on," although they will have no information concerning their personal circumstances, nor that of any particular persons (p. 197). It is the task of this convention to codify and institutionalize (e.g., in a Bill of Rights), the right of the citizens to equal liberty, as enunciated in the first principle of justice. The general laws, procedures,

and guarantees approved therein are to govern the work of the next stage: the *Legislative Stage*. At this stage, it is the task of the parties to enact laws and to designate officials and institutions to carry out specific economic and social policies. The distribution procedures established at this level are to be constrained by the second principle of justice (i.e., equal opportunity and the difference principle). Interestingly, the fact that the enforcement of the first principle is the responsibility of the Constitutional Convention and that the enactment of the second principle is the task of the subordinate legislative stage, reflects, in the political institutions, the rule of the priority of liberty.

In the final stage, the *Application Stage*, it is the task of judges and administrators to enforce and execute, and of citizens to follow, the principles, rules and laws set down in the higher stages (p. 199). Here, the veil of ignorance is totally lifted and thus all available knowledge admitted.

Just economic institutions. In a well-ordered society, the economic institutions distribute good according to the precepts of the principle of equal opportunity and the difference principle. As Rawls puts it:

The sum of transfers and benefits from essential public goods [is] arranged so as to enhance the expectations of the least favored consistent with the required savings and the maintenance of equal liberties. When the basic structure takes this form the distribution that results will be just (or at least not unjust) whatever it is. (p. 304)

In addition, the distributions of goods display "pure procedural justice," which is to say that "no attempt is made to define the just distribution of goods and services on the basis of information about the preferences and claims of particular individuals" (p. 304). Interestingly, Rawls believes that just economic distributions are possible both in free enterprise and collective (socialistic) systems of production (p. 271).¹⁶

Stability and just institutions. One factor that should weigh heavily in the deliberation of the original position is stability; that is, whether the principles chosen by the parties will generate institutions that they will wish to support and sustain. Another aspect of the same issue is the question of whether persons born and raised in a well-ordered society will "develop a desire to act in accordance with these principles and to do their part in institutions which exemplify them" (p. 177). Rawls believes that a society which exemplifies the principles of justice will clearly meet the test of stability, for under such conditions:

Each person's liberties are secured and there is a sense defined by the difference principle in which everyone is benefitted by social cooperation. Therefore we can explain the acceptance of the social system and the principles it satisfies by the psychological law that persons tend to love, cherish, and support whatever affirms their own good. Since everyone's good is affirmed, all acquire inclinations to uphold the scheme. (p. 177)

Among the personal goods affirmed in a well-ordered society are self-respect and the liberty to pursue one's own life plan. The concepts of "stability," the good of self-respect, and the rational plan of life,

have appeared from time to time in the foregoing sections. In the next section, these concepts will be given further scrutiny as we examine Rawls's "Full Theory of the Good."

26. "Goodness as Rationality"

"Right" and "good" to Rawls. The principles of justice draw the boundaries around the realm of the *right*. They do little, however, to indicate the content of the *good* – the landscape within the borders of this realm. To comprehend and to work toward the good for a society, an institution or a person, one must first understand the particular nature and circumstances of that society, institution, or person. However, such knowledge is forbidden in the original position. Now, according to Rawls's deontological perspective, nothing that violates the right can count as a *good* (p. 31). But, while this clearly tells us what is not good, it does little to illuminate our conception of what is good. For this, we need a closer look at human nature and then at the particular institution or person whose good we wish to understand. In the last of the three main parts of *A Theory of Justice* (titled "Ends"), Rawls explores the question of the good life of the individual, within the constraints of the *just* (or "right") life. In these final three chapters of his book (comprising almost 200 pages), Rawls treats such diverse topics as moral deliberation, moral sentiments (e.g., respect, guilt, shame), happiness, virtue and evil, moral development, and so on. Pervading all these chapters are the questions of *stability* and *congruence* (p. 395); namely: (a) does a well-ordered society produce members who will accept support and sustain the principles of justice upon which the society is based?; and (b) is a well-ordered society the sort of place where personal growth, satisfaction and flourishing can take place -- in a word, is *justice* a personal *good*?

But haven't we already examined Rawls's theory of the good -- namely, the index of *primary goods*? Not really. The primary goods are desired and needed by everyone, whatever else they may require. They fall under the so-called "thin theory of the good." Our present concern is with the goods denoted above by the phrase "whatever else." Rawls calls this the "Full Theory of the Good." He relates the two theories as follows:

We need what I have called the thin theory of the good to explain the rational preference for primary good and to explicate the notion of rationality underlying the choice of principles in the original position. This theory is necessary to support the requisite premises from which the principles of justice are derived. But looking ahead to other questions yet to be discussed, a more comprehensive account of the good is essential. Thus the definition of beneficent and supererogatory acts depends upon such a theory. So likewise does the definition of the moral worth of persons. This is the third main concept of ethics and we must find a place for it within the contract view. Eventually we shall have to consider whether being a good person is a good thing for that person, if not in general then under what conditions. (p. 397)

These final 200 pages of Rawls book are rich, suggestive and comprehensive -- both for moral philosophy in general and for the posterity question in particular. However, Rawls has little to say *explicitly* about the *good* of caring for posterity. As I have noted, his primary concern (expressed in

the first two parts of the book) has been for the *rightness* of making just provision for future generations. Even so, as we shall discover in succeeding chapters, Rawls's "Full Theory of the Good," does, in fact, bear significant implications for the question of the duty to posterity.

Goodness as rationality. According to Rawls, the individual's good is achieved through the successful pursuit of a "rational plan of life." What does he mean by this? He explains that:

A person's good is determined by what is for him the most rational long-term plan of life given reasonably favorable circumstances. A man is happy when he is more or less successfully in the way of carrying out this plan. To put it briefly, the good is the satisfaction of rational desire. We are to suppose, then, that each individual has a rational plan of life drawn up subject to the conditions that confront him. This plan is designed to permit the harmonious satisfaction of his interests. It schedules activities so that various desires can be fulfilled without interference. It is arrived at by rejecting other plans that are either less likely to succeed or do not provide for such an inclusive attainment of aims. Given the alternatives available, a rational plan is one which cannot be improved upon; there is no other plan which, taking everything into account, would be preferable. (pp. 92-93)

In connection with "life plans," the term "rationality" is given two complementary interpretations:

A person's plan of life is rational if, and only if, (1) it is one of the plans that is consistent with the principles of rational choice when these are applied to all the relevant features of his situation, and (2) it is that plan among those meeting this condition which would be chosen by him with full deliberative rationality, that is, with full awareness of the consequences.¹⁷ . . . [A] person's interests and aims are rational if, and only if, they are to be encouraged and provided for by the plan that is rational for him We criticize someone's plan, then, by showing either that it violates the principles of rational choice, or that it is not the plan that he would pursue were he to assess his prospects with care in the light of a full knowledge of his situation. (pp. 408-409)

It is important to note that this definition of a person's good as "the satisfaction of rational desire" is purely formal. We cannot, from this definition alone, determine "what sorts of ends these plans are likely to encourage. In order to draw conclusions about these ends, it is necessary to take note of certain general facts". These "general facts" include data of human desires, needs, capacities, and abilities (p. 424). In addition, to determine his own particular life plan, each person must take into account aspects of his background and circumstances that are unique to him.

With this definition of "good as rationality" at hand, we are prepared to explicate such morally significant concepts as *happiness*, *the good object*, *the good person* (i.e., of moral worth), and *the good* (or beneficent) *act*. First, *happiness* is seen to have two aspects:

. . . one is the successful execution of a rational plan . . . which a person strives to realize, the other is his state of mind, his sure confidence supported by good reasons that his success will

endure. Being happy involves both a certain achievement in action and a rational assurance about the outcome. (p. 549)

Second, an *object* is good (for someone) if it "has the properties that it is rational for someone with a rational plan of life to want" (p. 399). Third, a *good person*, or a person of moral worth, "is someone who has to a higher degree than the average the broadly based features of moral character that it is rational for the persons in the original position to want in one another." In other words, "a good person has the features of moral character that it is rational for members of a well-ordered society to want in their associates" (p. 437). Finally, a *good act* (or beneficent act) can be defined as an act: ". . . which we are at liberty to do or not to do, that is, no requirements of natural duty or obligation constrain us either to do it or not to do it, and which advances and is intended to advance another's good (his rational plan)." (p. 438)

The good and the right contrasted. It should now be clear, from the foregoing account, that the *good* for man is quite distinct from his *duty* -- that is, from what is *right* (i.e., in accordance with the principles of justice). Of course, these categories of morality, while distinct, are, of necessity, never *in conflict*. Rawls describes the distinction with admirable clarity. *First*, he reminds us, the right follows from the conditions in the original position, while the (full) *good* does not. He writes: "Whereas the principles of justice (and the principles of right generally) are those that would be chosen in the original position, the principles of rational choice and the criteria of deliberative rationality are not chosen at all." (p. 446)

He continues by pointing out that, while the principles of justice must be agreed to unanimously in the original position, this is not the case with the theory of the good. There is, he says: ". . . no necessity for an agreement upon the principles of choice. Since each person is free to plan his life as he pleases (so long as his intentions are consistent with the principles of justice), unanimity concerning the standards of rationality is not required." (p. 447)

A *second* distinction (implicit in the above account of the first) is that the principles of right are invariable, while personal conceptions of the good may vary. Says Rawls:

It is, in general, a good thing that individuals' conceptions of their good should differ in significant ways, whereas this is not so for conceptions of right. In a well-ordered society citizens hold the same principles of right and they try to reach the same judgment in particular cases. These principles are to establish a final ordering among the conflicting claims that persons make upon one another. . . . On the other hand, individuals find their good in different ways, and many things may be good for one person that would not be good for another. Moreover, there is no urgency to reach a publicly accepted judgment as to what is the good of particular individuals. The reasons that make such an agreement necessary in questions of justice do not obtain for judgments of value. (pp. 447-448)

The *third* difference is that the principles of justice (right) are chosen behind the veil of ignorance, while the evaluation of one's good requires full knowledge. Thus, says Rawls:

Not only must the principles of justice be chosen in the absence of certain kinds of particular information, but when these principles are used in designing constitutions and basic social arrangements, and in deciding between laws and policies, we are subject to similar although not as strict limitations An individual's conception of his good, on the other hand, is to be adjusted from the start to his particular situation. A rational plan of life takes into account our special abilities, interests, and circumstances, and therefore it quite properly depends upon our social position and natural assets. (p. 449)

Underlying all these distinctions, of course, is the fundamental rule: *the right is prior to the good*, and thus sets limits upon what can be counted as good.

Moral psychology and the sense of justice. The "Sense of Justice" is described by Rawls as "a settled disposition to adopt and to want to act from the moral point of view, insofar at least as the principles of justice define it" (p. 491). This sense entails both a willingness to confine one's life-plan within the constraints of justice, and a desire to support just institutions where they exist and to work toward their establishment where they do not exist. A sense of justice is to be found (all too infrequently) among members of all varieties of civilized society, however imperfect. The sense is manifested in acts of civil disobedience and conscientious objection, in political debate, and in such political documents as the United States Constitution and the Bill of Rights.

In a well-ordered society, a sense of justice is a virtually universal endowment to those fortunate enough to grow up in such a society. The sense is acquired in three stages, described in Rawls's "Laws of Moral Development." The first law ("the Morality of Authority") states:

Given that family institutions are just, and that the parents love the child and manifestly express their love by caring for his good, then the child, recognizing their love of him, comes to love them. (p. 490)

The second law ("the Morality of Association") asserts:

Given that a person's capacity for fellow feeling has been realized by acquiring attachments in accordance with the first law, and given that a social arrangement is just and publicly known by all to be just, then this person develops ties of friendly feeling and trust toward others in the association as they with evident intention comply with their duties and obligations, and live up to the ideals of their station. (p. 490)

Finally, the third law ("the Morality of Principles") reads:

Given that a person's capacity for fellow feeling has been realized by his forming attachments in accordance with the first two laws, and given that a society's institutions are just and are publicly known by all to be just, then this person acquires the corresponding sense of justice as he recognizes that he and those for whom he cares are the beneficiaries of these arrangements. (p. 491)

Rawls is describing here a process that is reminiscent of Socrates' account of "the ladder of *love*" in Plato's *Symposium*. In Rawls's version, the child first develops a love and a loyalty to those most immediately and conveniently present and caring -- his parents. The loyalty is extended, to relatives and friends, and then to such abstractions as associations and institutions to which one's acquaintances (and oneself) belong. Finally, the loyalty attaches to the most abstract of entities, *ideas* and *principles*.¹⁸ A dramatic moral crisis, such as the recent Watergate Scandal, often illustrates the conflict between these three stages of morality. In the Watergate affair, some officials were motivated by their loyalty to a person, i.e., Mr. Nixon. Others were moved by their loyalty to an institution, i.e., the Presidency. Still others acted in accordance with their duty to uphold the general principle of equal justice under the law.

Those who possess a sense of justice must pay a price for it. The price is the liability to suffer *guilt* and *shame* when one's conduct fails to meet the standards (of *right*) implied by the principles of justice, or the virtues and excellences (i.e., of *goodness*) consistent with one's personal plan of life. In general, says Rawls, *guilt* on the part of the agent, and *resentment* and *indignation* on the part of those affected by the agent, "invoke the concept of right." On the other hand, *shame* on the part of the agent, and *contempt* and *derision* from others, "invoke the concept of goodness"^a (p. 484). Rawls explains the difference in this manner:

While the principles of right and justice are used to describe the actions disposing us to feel both moral shame and guilt, the perspective is different in each case. In the one [guilt] we focus on the infringement of the just claims of others and the injury we have done to them, and on their probable resentment or indignation should they discover our deed. Whereas in the other [shame] we are struck by the loss to our self-esteem and our inability to carry out our aims; we sense the diminishment of self from our anxiety about the lesser respect that others may have for us and from our disappointment with ourselves for failing to live up to our ideals. Moral shame and guilt, it is clear, both involve our relations to others, and each is an expression of our acceptance of the first principles of right and justice. Nevertheless, these emotions occur within different points of view, our circumstances being seen in contrasting ways. (p. 446)

While the liability of bearing the weight of guilt and shame may be burdensome, and consequences of avoiding this liability are far worse. In one of the most moving and profound passages in his book, Rawls explains:

A person who lacks a sense of justice, and who would never act as justice requires except as self-interest and expediency prompt, not only is without ties of friendship, affection, and mutual trust, but is incapable of experiencing resentment and indignation. He lacks certain

^aI have a somewhat different sense of "guilt" and "shame." *Shame* focuses upon one's concern about the opinion of *others* about one's moral worth. *Guilt* focuses upon one's sense of diminished moral self-worth.

natural attitudes and moral feelings of a particularly elementary kind. Put another way, one who lacks a sense of justice lacks certain fundamental attitudes and capacities included under the notion of humanity. Now the moral feelings [guilt and shame] are admittedly unpleasant, in some extended sense of unpleasantness; but there is no way for us to avoid a liability to them without disfiguring ourselves. This liability is the price of love and trust, of friendship and affection, and of a devotion to institutions and traditions from which we have benefited and which serve the general interests of mankind. Further, assuming that persons are possessed of interests and aspirations of their own, and that they are prepared in the pursuit of their own ends and ideals to press their claims on one another -- that is, so long as the conditions giving rise to questions of justice obtain among them -- it is inevitable that, given temptation and passion, this liability will be realized. And since being moved by ends and ideals of excellence implies a liability to humiliation and shame implies a lack of such ends and ideals, one can say of shame and humiliation also that they are part of the notion of humanity. Now the fact that one who lacks a sense of justice, and thereby a liability to guilt, lacks certain fundamental attitudes and capacities is not to be taken as a reason for acting as justice dictates. But it has this significance: by understanding what it would be like not to have a sense of justice -- that it would be to lack part of our humanity too -- we are led to accept our having this sentiment. (pp. 488-489) (See also Rawls's early paper, "The Sense of Justice" 1963, p. 282.)

Later in this dissertation, as I have occasion to draw motivations for caring for future generations from Rawls's theory of the good and elsewhere, I will return to the ideas expressed in this significant account of the sense of justice -- and of the terrible deprivations of those without it. (See §§43, 45, below.)

"The Kantian interpretation": A critical note. The reader familiar with Rawls's treatment of the sense of justice may wonder why I have failed heretofore to mention his "Kantian Interpretation" of justice as fairness. The omission has been deliberate, as I will explain shortly. But first, I should briefly describe the "Kantian interpretation."

Rawls believes that his principles of justice can be compared to the categorical imperative and other aspects of Kant's moral philosophy. Accordingly, a desire and disposition to act in accordance with these principles (i.e., a "sense of justice") displays a full fruition of the human capacity for autonomy, equality, and rationality (p. 253). Says Rawls:

The desire to act justly derives in part from the desire to express most fully what we are or can be, namely free and equal rational beings with a liberty to choose Liberty is acting in accordance with a law that we give to ourselves. And this leads not to a morality of austere command but to an ethic of mutual respect and self-esteem. (p. 256)

Thus, "The original position may be viewed . . . as a procedural interpretation of Kant's conception of autonomy and the categorical imperative" (p. 256).

Oliver Johnson (1974) rejects Rawls's Kantian interpretation, arguing that "the conception that [Rawls] has of man's nature as a moral being is basically opposed to, rather than consonant with, that held by Kant" (p. 58). Since I cannot report Johnson's full argument here, one aspect will have to suffice. Rawls, says Johnson, believes that "the decision the individual makes in choosing the principles of justice is rational insofar as it serves his long-range best interests. For him, thus, reason is instrumental; its function is to select means that will lead to ends determined by desire" (p. 64). However, to Kant:

The moral and proper function of reason is to produce a will good in itself. That Kant's conception of moral or practical reason is not only different from, but opposed to, the account offered by Rawls is apparent. For Rawls's paradigm example of the moral use of reason Kant would deny to have anything to do with morality at all. And Kant's description of the proper moral function of reason is nowhere echoed in Rawls's theory. Rather than being consonant with each other, these two conceptions of the role of reason in the moral life stand, in relation to each other, very near the limits of incompatibility. No 'Kantian interpretation' is remotely possible. (pp. 65-66).

I find Johnson's argument here, and elsewhere, to be convincing and conclusive. Accordingly, I have generally chosen, in this dissertation, to bypass Rawls's "Kantian interpretation." A strong case can be made, I believe, that an effective sense of justice displays moral rationality and autonomy. But this "rationality" and "autonomy" are "Kantian" in only the most superficial sense, and the deeper grounds and implications of freedom and reason, in Rawls and in Kant, diverge radically. I cannot in this space, delve into these "deeper grounds." Suffice it to say that, to Rawls, *reason* is but one of several factors involved in the derivation of the principles of right, while, for Kant, the right is exemplified in the willingness to act "categorically" according to the dictates of pure reason (i.e., apart from "hypothetical" calculations of results). (I will not deal here with Johnson's other objections to "the Kantian interpretation".)

Where does this leave Rawls's theory? Essentially untouched, I believe. While Rawls has devised the "Kantian interpretation" as a means of *illuminating* justice as fairness, it plays no significant role in the *justification* of his theory. Thus, while Rawls's scholarship might, in this case, be questioned (a rare and uncharacteristic lapse:), the integrity of his system has been little compromised.¹⁹

The Aristotelian Principle. Rawls believes that human activity, both individual and communal, tends toward complexity and comprehensiveness. This, he contends, is a basic fact of human nature which must be taken into account in assessing the goodness of personal life plans, or of a social order. This tendency, which Rawls calls "the Aristotelian Principle," indicates that:

Other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and that this enjoyment increases the more the capacity is realized, or the greater its complexity. A person takes pleasure in doing something as he becomes more

proficient at it, and of the two activities which he performs equally well, he prefers that the one that calls upon the greater number of more subtle and intricate discriminations. (p. 414)

The Aristotelian Principle, says Rawls, is evident in countless events of everyday life; notably in the play of children, in the arts, and in personal hobbies. All this indicates that "human beings enjoy the greater variety of experience, they take pleasure in the novelty and surprises and the occasions for ingenuity and invention that such activities provide" (p. 431). People also seem to prefer the complex because of "the pleasures of anticipation and surprise" that they evoke, and they are attracted by the opportunity to display, in skillful activity, "individual style and personal expression" (p. 427).

If Rawls has correctly described the Aristotelian Principle as "a deep psychological fact," (p. 432) then it surely deserves a prominent place in the "rational life plan" which determines a person's good. Furthermore, as an individual exhibits the Aristotelian Principle, in his professional work, his personal associations, and other aspects of his life, he will enjoy the admiration of others and thus enhance his self-esteem (a primary good). The principle yields obvious social advantages as well which are manifested in the progress of the arts, the sciences, technology, and other aspects of civilized life that exhibit skill, complexity, and comprehensiveness (p. 429).

The good community. As we now know, in Justice as Fairness, the principles of justice are derived from the original position -- a conceptual model devised to exhibit the results of rational deliberation by self-serving ("mutually disinterested") *individuals*. In addition, Rawls's account of the good employs such concepts as "rational life plans," "the sense of justice," "the Aristotelian Principle," all of which are treated as accomplishments by the *individual*. However, all this should not lead us to suspect that Rawls has overlooked the necessary contribution of the *community* to the attainment of human well-being and fulfillment. Indeed, he insists that "human flourishing" can best take place in the context of a well-ordered community, regulated by the principles of justice. Rawls writes:

The collective activity of justice is the preeminent form of human flourishing. For given favorable conditions, it is by maintaining these public arrangements that persons best express their nature and achieve the widest regulative excellences of which each is capable. At the same time just institutions allow for and encourage the diverse internal life of associations in which individuals realize their more particular aims. Thus the public realization of justice is a value of community. (p. 528)

Why is this? Why is the community necessary for human flourishing? This is so simply because collective association offers the individual security and resources that he could never attain by himself. For example, while in a flourishing society one can appreciate many arts; he can master only one, if that. While he can benefit from the findings of many sciences and the applications of many crafts, he can become the expert of only one, if that. In short, a community offers the individual the opportunity to benefit, in his single lifetime, from the experience and contributions of countless others. The English economist, L. T. Hobhouse, expressed the point quite vividly when he wrote:

The organizer of industry who thinks he has 'made' himself and his business has found a whole social system ready to his hand in skilled workers, machinery, a market, peace and order -- a vast apparatus and a pervasive atmosphere, the joint creation of millions of men and scores of generations. Take away the whole social factor and we have not Robinson Crusoe, with his salvage from the wreck and his acquired knowledge, but the native savage living on roots, berries and vermin.²⁰

Rawls concurs, as he writes (somewhat less elegantly): "[Because] our potentialities and inclinations far surpass what can be expressed in any one life, we depend upon the cooperative endeavors of others not only for the means of well being but to bring to fruition our latent powers" (p. 571).

However, as Hobhouse clearly indicates, the advantages of community life are not entirely derived from one's contemporaries. There is also a rich legacy from the past, a legacy of "systems of knowledge and belief," of "recognized tendencies for doing things, and elaborate styles of feeling and expression" (Rawls, 1971, p. 526). As the members of the community perceive, share, and benefit from their common political, artistic, scientific, and religious traditions, they may also perceive themselves as participants in an ongoing historical process in which they are aiming toward the eventual achievement of "shared final ends" (p. 526). The sense of community, then, may well be extended to encompass one's ancestors and one's posterity. One's own self-esteem may well be affected by the extent to which he contributes his personal time, talents and career to social and cultural projects and goals which transcend himself. (I will have much more to say about "self transcendence" and the provision for posterity in Chapter VII, below.)

However, says Rawls, if all these worthy personal dividends and collective accomplishments are to be realized through the flourishing of "the good community":

We must acknowledge the principles of its regulative conception, and this means that we must affirm our sentiment of justice. To appreciate something as ours, we must have a certain allegiance to it. What binds a society's efforts into one social union is the mutual recognition and acceptance of the principles of justice; it is this general affirmation which extends the ties of identification over the whole community and permits the Aristotelian Principle to have its wider effect. (p. 571)

The congruence of the right and the good. In this chapter we have examined the derivation of the principle of justice (the right) from the original position. We have also discussed the nature and development of the sense of justice in the individual. Somewhat separately, we have considered Rawls's concept of personal good as "the satisfaction of rational life plans." These issues may now be drawn together as we ask, with Rawls, whether the Right and the Good are *congruent*. Rawls poses the question of congruence in this manner: In a well ordered society, does "a persons is rational plan of life affirm and support his sense of justice?" And the converse question: "Do the various desiderata of a well ordered society and . . . its just arrangements [i.e., institutions] contribute to the good of its members?"²¹ (p. 513). In both cases, Rawls believes that the answer is affirmative. Most of its supporting arguments are familiar to us by now.

Why, then, should we believe that justice serves one's personal good? In the first place, as we have seen, the principles of justice foster and support self-respect which, in turn, enhances social solidarity (p. 178). (See pp. 101-2, above). Rawls is quite explicit about the importance of self-respect to a person's good: "It is clearly rational for men to secure their self-respect. A sense of their own worth is necessary if they are to pursue their conception of the good with zest and to delight in its fulfillment. Self-respect is not so much a part of any rational plan of life as the sense that one's plan is worth carrying out" (p. 178). Secondly, a just ("well ordered") society is best equipped to facilitate the "human flourishing" that was just noted; namely, the communal sharing, enjoyment and advancement of individual expertise, talents and accomplishments. Finally, the sense of justice exacts psychological penalties upon those whose life-plans violate the principles of right. Such individuals are liable to suffer feelings of *guilt* and *shame*, and be regarded by other with *indignation* and *contempt*.

And what reason is there to believe that the *good* upholds the *right* – that a rational plan of life supports the principles of justice? To begin, a person capable of ordering his life with "rational deliberation" should also be able to perceive that the principles of justice are such that he himself would choose them under explicitly fair conditions of deliberation; namely, the conditions of the original position. Furthermore, according to the three laws of moral development (see pp. 116-7, above), as a person matures, his good is naturally extended to include the well-being of his parents, his family, his friends, and his associates, and eventually associations and institutions. (With full development, his loyalty reaches to ideas and principles -- but that is not the point at issue right now.) To desire the well-being of those we love includes the desire to be *fair* to them; i.e., to act justly (p. 570). Of course, one might attempt to be fair only to those closest to him, and to act hypocritically and deceptively toward others. However, in a closeknit society, it is not easy to "select who is to be injured by our unfairness," or to predict whose loyalty we may need or whose friendship we may cherish in the future. Thus, even from the limited standpoint of personal prudence, the best policy may well be to extend just treatment to as many persons, associations, and institutions as possible. In the best of circumstances, which is to say in a well-ordered society, one would apply the principles of justice *universally* (pp. 570-571).

If Rawls's arguments for the congruence of the right and the good are sound, then both aspects of morality would be coherently combined in the virtuous person. Such a person, says Rawls, would display two complementary capacities:

One for a conception of the good, the other for a sense of justice. When realized, the first is expressed by a rational plan of life, the second by a regulative desire to act upon certain principles of right. Thus a moral person is a subject with ends he has chosen, and his fundamental preference is for conditions that enable him to frame a mode of life that expresses his nature as a free and equal rational being as fully as circumstances permit. Now the unity of the person is manifest in the coherence of his plan, this unity being founded on the higher order desire to follow, in ways consistent with his sense of right and justice, the principles of rational choice. (p. 561)

A community of such virtuous persons would, of course, be a *stable* community in that the members thereof would acknowledge that it is to the advantage of *each* that all honor the principles of justice (p. 576). Thus, in summarizing his argument for congruence, Rawls concludes:

First that, in a well-ordered society, being a good person (and in particular having an effective sense of justice) is indeed a good for that person; and second that this form of society is a good society Thus a well-ordered society satisfies the principles of justice which are collectively rational from the perspective of the original position; and from the standpoint of the individual, the desire to affirm the public conception of justice as regulative of one's plan of life accords with the principles of rational choice. These conclusions support the values of community, and in reaching them my account of justice as fairness is completed. (p. 577)

And so too is my presentation of justice as fairness. All that remains, in this chapter, is a brief account of Rawl's response to his rivals, and some concluding meta-ethical comments on his theory.

27. Rival Theories: Perfectionism and Utilitarianism

In the original position, the parties have the opportunity to examine and assess a variety of moral theories for possible adoption as "rules of their association." Among the alternatives are justice as fairness, "perfectionism," various types of utilitarianism, intuitionism, egoism, and "state of nature anarchy" (the last being a "decision of no-decision") (Rawls, §21). Of course, Rawls believes that the parties would choose justice as fairness: indeed, the primary point of his book is to demonstrate the validity of this belief. Of all the alternatives considered in the original position, Rawls takes perfectionism and utilitarianism the most seriously. We will do likewise. A review of his criticisms of these two moral theories will serve to place the boundaries of justice as fairness into sharper focus.

Perfectionism. The lives of ordinary mortals, says Nietzsche, have meaning only in the service of the superior. He writes: "Mankind must work continually to produce individual great human beings -- this and nothing else is the task. . . for the question is this: how can your life, the individual life, retain the highest value, the deepest significance? . . . Only by your living for the good of the rarest and most valuable specimens" (Nietzsche, quoted by Rawls, p. 325n).²² This is a radical version of "the principle of perfection" which would surely be rejected in the original position, since there would be virtually no chance for any of the parties to benefit therefrom. A moderate form of the perfection principle would read:

The greater happiness of the less fortunate does not in general justify curtailing the expenditures required to preserve cultural values. These forms of life have greater intrinsic worth than the lesser pleasures, however widely the latter are enjoyed. Under normal conditions a certain minimum of social resources must be kept aside to advance the ends of perfection. The only exception is when these claims clash with the demands of the basic needs. Thus given improving circumstances, the principle of perfection acquires an increasing weight relative to a greater satisfaction of desire. (p. 326)

Rawls believes that this too would be rejected. Since Rawls's refutation of moderate perfectionism entails a refutation of the radical kind as well, we will confine our attention to his criticism of the former.

Why, then, would even a moderate principle of perfectionism not be acceptable in the original position as a component "rule of association" in actual life? The problem is that the traits to be maximized are not *primary* goods, which all rational agents would desire and need. In fact, says Rawls, "criteria of excellence are imprecise as political principles, and their application to public questions is bound to be unsettled and idiosyncratic, however reasonably they may be invoked and accepted within narrower traditions and communities of thought" (pp. 330-331). "The excellences," therefore, belong to personal conceptions of the good which, as we have seen, are barred from the original position by the veil of ignorance. It follows that no decision could be made to maximize goods that were not identifiable in the original position.

However, the fact that the maximization of cultural or artistic excellences is *not* among the principles of justice by no means bars these perfections from the list of *personal* goods. In a well-ordered society, persons are quite free to pursue, develop, enjoy and support these enterprises, either individually or collectively. But they may not force others, who do not share their tastes, to do so.

While justice as fairness allows that in a well-ordered society the values of excellence are recognized, the human perfections are to be pursued within the limits of the principle of free association. Persons join together to further their cultural and artistic interests in the same way that they form religious communities. They do not use the coercive apparatus of the state to win for themselves a greater liberty or larger distributive shares on the grounds that their activities are of more intrinsic value. Perfectionism is denied as a political principle. (pp. 328-329)

Utilitarianism. By far the largest share of Rawls's critical remarks are directed against utilitarianism: both the "classical" conception (which stresses the maximization of *total* utility), and the more recent revisions (which argue for the maximization of *average* utility) (Rawls, p. viii).^a Many commentators have suggested that Rawls's most significant contribution might be his reintroduction of contract theory as a deontological rival to the utilitarianism which has so dominated the thinking of English-speaking moral philosophers this past century (p. vii). (See also, Cohen, 1972; Feinberg, 1973, p. 263; Hampshire, 1972; and Lyons, 1972). The influence of utilitarianism, and the consequent significance of Rawls's rebuttal, extends beyond moral philosophy to economics (Arrow, 1973, p. 246), political science (Hampshire, 1972, p. 75) and the so-called "policy sciences" (Tribe, 1972).

^aIt is important to note that the issue of *average v. total* utility arises only with regard to provision for the future – particularly with the consideration of the size of a future population. Given a constant population (e.g., in the case of policies confined to time-present), average utility is completely a function of total utility.

But for Rawls's constant reiteration of his opposition to utilitarianism, the casual reader might be struck by the similarities in the theories. In many aspects, justice as fairness *sounds* utilitarian; for instance, in the discussions of the "distribution of goods," or "maximizing prospects." Nowhere, however, does Rawls suggest that the criterion of the right is the maximization of *total* or *average* good. As we have seen, Rawls's theory is quite at odds with these generalized maximization principles. And, most fundamentally, Rawls's theory defines the *right* as independent of, and absolutely prior to, the *good*. This essential insistence upon the independence and priority of the *right*, places justice as fairness in irrevocable opposition to utilitarianism.

Rawls's confrontations with utilitarianism are so numerous and extensive throughout his book that more than an entire column of space is required simply to list them in his index (pp. 606-607). Moreover, as just suggested, numerous papers have been written in response to the philosophical disputes between justice as fairness and utilitarianism. Obviously, I cannot in this brief space begin to do justice to these issues. However, a brief statement of three major points of contention may be illuminating. *First* of all, (a) "utilitarianism does not take seriously the distinction between persons" (p. 27). To the utilitarians, the amount (total or average) of good distributed to *all* is what matters -- not the rules of distribution to *each*. Rawls is pre-eminently concerned with the (*right*) principles of distribution. *Secondly*, to the utilitarian, the optimum society is one which distributes resources with the greatest efficiency; that is, in a manner productive of the greatest benefit (total or average) as this distribution might be assessed by an hypothetical, omniscient "ideal observer." Justice as fairness, on the other hand, conceives the just society as a scheme of association that would be devised by an assembly of rational egoists, for their reciprocal advantage, under conditions that all would agree are fair (p. 33). Because of this greater emphasis upon the individual and his equal access to advantages, Rawls believes that his conception would be preferred by the parties in the original position (p. 14). *Finally*, Rawls charges that utilitarianism asks too much of some individuals; namely, that they "should forego advantages for the sake of the greater good of the whole" (p. 177). Justice as fairness, however, assures equal liberty to all and stipulates that unequal distributions of (primary) goods are allowable only if the least endowed gain thereby. It follows that Rawls's "well ordered society" is far more stable than the society based upon utilitarian principles. Rawls's explanation of this point betrays a touch of sarcasm:

It is evident . . . why utilitarians should stress the role of sympathy in moral learning and the central place of benevolence among the moral virtues. Their conception of justice is threatened with instability unless sympathy and benevolence can be widely and intensely cultivated. Looking at the question from the standpoint of the original position, the parties recognize that it would be highly unwise if not irrational to choose principles which may have consequences so extreme that they could not accept them in practice. (p. 178)

All too abruptly, we must now close this account of Rawls's criticism of utilitarianism. We will occasionally return to this contest as it applies to the question of just provision for future generations.

28. Justice as Fairness: A Meta-ethical Postscript

I will not attempt here to summarize a chapter that is itself a summary. However, a few closing reflections concerning Rawls's impressive contribution to contemporary moral philosophy may lend some perspective to this lengthy and wide-ranging chapter.

My primary objective in this chapter has been to present the essential elements of Rawls's theory of justice and at least to suggest its rich systematic integrity and coherence. Just this objective has entailed much more than ample material to fill this lengthy chapter. Accordingly, I have had little opportunity to present Rawls's supporting arguments, much less to review the many published critical responses or to develop many of my own. I have tried to express Rawls's position with sympathy and in a favorable light; as if, that is to say, I were generally convinced by his presentation and converted to his position.

In point of fact, while I am profoundly impressed by the breadth of Rawls's knowledge, the depth of his scholarship, and the elegance of his system, I remain sceptical about many particular aspects of his theory. For the most part, my critical responses to justice as fairness must remain unstated. However, as promised, I will later attempt careful, detailed and extensive analyses of one aspect of Rawls's thought: the question of "justice between generations."

To close this chapter, I would like to list what I believe to be some of the most significant features, and, I would hope, the most enduring aspect of Rawls's theory of justice.

Justice as fairness is a substantive and normative theory. Contrary to fashionable philosophical practice, Rawls is not content simply to *analyze* moral discourse, or to *describe* moral deliberation. He believes that "it is obviously impossible to develop a substantive theory of justice founded solely on truths of logic and definition" (p. 51); and "a substantive theory" is precisely what he has developed. This theory both describes and *prescribes* just principles and virtuous life-plans. Rawls, in short, is attempting to revive, and proposing to answer, such fundamental ethical questions as: "What is a just community?", "What are the rights and duties of the members of such a community?", "What is a good life?"

Justice as fairness is drawn from, and is relevant to, ordinary practical moral experience and reflection. Ordinary experience, however, is by no means the exclusive source, for "considered moral judgments," in Rawls's theory, are refined, formalized, and synthesized into general principles which, through "reflective equilibrium," affect the outcome of practical moral judgments. Justice as fairness begins as an "attempt to describe our moral capacity," (p. 46) and continues as a means of informing, refining, and guiding this capacity. The close coordination of Rawls's theory with practical moral experience is, I believe, one of its most distinctive and commendable features. (For a sample of the practical implications of justice as fairness, see Rawls's discussion of civil disobedience in §§53-59 of his book.)

Justice as fairness claims only to be an approximation. The scope of Rawls's theory is broad, the answers he seeks are wide-ranging, and the enterprise that he has projected is ambitious. In contrast,

his claims of success are modest. As he perceives the present condition of moral philosophy (including his own), he finds that:

[The] 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. (p. 52)

Rawls makes no claim to have written the final moral philosophy. He does believe, however, that, in view of the considerable advances in scientific knowledge and philosophical method since the days of Locke, Hume and Kant, the traditional theory of the social contract should be updated and given a careful and scrupulous hearing. The critical response that has followed the publication of Rawls's book indicates that he has eminently succeeded in reviving this debate among moral philosophers, and beyond.

Justice as fairness is fact-contingent and thus subject to change and growth. This feature, which is evident throughout the book, is clearly explicated in the final section (§87), where Rawls contrasts his mode of theoretical justification with the "Cartesian" and "Naturalistic" approaches. The "Cartesian method," Rawls explains, "presumes that first principles can be seen to be true, even necessarily so; deductive reasoning then transfers this conviction from premises to conclusion" (p. 578). The "naturalistic approach," he says:

Is to introduce definitions of moral concepts in terms of presumptively non-moral ones, and then to show by accepted procedures of common sense and the sciences that the statements thus paired with the asserted moral judgments are true. Although on this view the first principles of ethics are not self-evident, the justification of moral convictions poses no special difficulties. They can be established, granting the definitions, in the same fashion as other statements about the world. (p. 578)

Rawls rejects both of these modes of justifying ethical judgments and principles. (Also, as I have just noted above, he believes that the analysis of concepts and meanings has no "special place in a theory of justice" (p. 51).) In short, "clear and distinct" *a priori* ideas, definitions, conceptual analysis -- all of these may be important *ingredients* of a theory of justice, but none alone can suffice to justify a theory. Among the additional criteria for justification are (a) *coherence*, and (b) entailments with, and from, factual data (p. 159). Since I have already examined the coherence criterion (§22, above), I will turn immediately to the other matter of "fact- contingency."

In what sense, then, is justice as fairness dependent not on *a priori* intuitions or definitions but on hard empirical *facts*. In the first place, as we have noted several times, in Rawls's theory, the principles of justice are in "reflective equilibrium" with "considered moral judgment" made in the

circumstances of ordinary life. Of course, the truth, validity, soundness, or justice of these judgments are not immediate data of experience, and Rawls does not claim this. However,, that certain persons make such and such moral judgments, and that these judgments are rationally "considered" -- these *are* fact claims that can be tested by such methods as interviews, questionnaires, case studies, etc. In this sense, then, "considered moral judgments" are genuinely factual ingredients of Rawls's theory.

In addition, we will recall that the parties of the original position are permitted to have access to all *general* knowledge that may bear upon the problem of determining the principles of justice. Throughout the book, Rawls suggests items of general information that might have significant bearing upon the deliberations in the original position. Since we have already covered this ground, a simple enumeration of these presumed facts will suffice. They include: (a) the list of desiderata that *all* would want, whatever else each *might* want (i.e., "the index of primary goods"); (b) the presumption that the least advantaged do in fact gain from larger distributions to others (the "trickle-down theory," which is an aspect of the difference principle); (c) the belief that persons generally tend to prefer complex and inclusive activities and skills to those that are simple and confined ("the Aristotelian Principle"); (d) the belief that the sense of justice evolves through three general stages of moral development; and finally, (e) the assertion that persons will generally support, affirm, and seek to perpetuate a social order that operates according to the principles of justice (a "well ordered society"), and that such a social order will usually facilitate and sustain a person's pursuit of his rational life-plan (the stability and congruence factors). All of these claims. and many others in *A Theory of Justice* are, I believe, amenable to empirical examination -- e.g., by anthropologists, economists and psychologists. Indeed, even a casual reading of a few of the responses to Rawls's book supplies ready illustration of the factual entailments of justice as fairness.

Justice as fairness is a theory. It is not simply a generalization from particular moral judgments or intuitions. Like such scientific theories as the kinetic theory of gases and the theory of organic evolution, justice as fairness provides a *model*, a *vocabulary*, and *rules of interpretation* by which a body of data or "class of facts" might be coordinated and structured. Rawls identifies this "class of facts against which conjectured [i.e., theoretical] principles can be checked" as "our considered judgments in reflective equilibrium" (p. 51). (Note the comparison made between justice as fairness and linguistic theory, note 4, p. 87, above.) Of course, we must not make so much of the similarities with science that we lose sight of the differences. Scientific theories seek to facilitate empirical prediction and verification, to unify disparate laws and hypotheses, and to suggest further avenues of research and inquiry. Normative *moral* theory remains fundamentally *practical*; that is, it is concerned with articulating rules for directing the conduct of life.

Justice as fairness is timely. For at least the last quarter century, Anglo-American moral philosophy has been dominated by the analytical approach and temperament. Now, at last, there is a decided resumption of interest by philosophers in substantive moral issues. Of course, while Rawls's work *reflects* this interest, and has contributed to its intensification, he is by no means the *cause* of it. Moreover, normative ethics has not merely been dormant and unaffected these past few decades waiting, as it were, for the analysts' dominance to fade and pass. On the contrary, the recent emphasis upon critical ethics has left a lasting and beneficial impact upon normative moral philosophy. Due to

recent explorations in meta-ethics, simplistic approaches, such as the "Cartesian" and "naturalistic" methods described above, will in the future receive more careful and capable scrutiny. Traditional moral theories, such as the contract theory, will require considerable elaboration and refinement if they are to attract and sustain the attention and adherence of future philosophers. Indeed, as I have previously indicated, *A Theory of Justice* is an eminent example of a traditional theory (i.e., contractualism) which has reflected, responded to, and in significant instances incorporated recent developments in critical philosophy.

In short, I am suggesting that philosophy has passed through a critical phase of its history, similar, in many ways, to those prompted in the past by the thought of such men as Socrates, of William of Occam, of Hume and of Kant. As a result of this season of analysis, many of the earlier speculative errors will be less likely to appear again in the mainstream of philosophical thought. For this, the philosophical profession should be duly grateful. However, with these new and important critical tools now at their disposal, it may well be time for more (but surely not all) moral philosophers to turn once more to the substantive issues.²³

Unquestionably, this is an age which is dominated by profound moral issues: such issues as war and peace, racial justice and equality, environmental renewal, personal integrity versus alienation, the human control of technology, to name but a few. These issues are inescapable and will be responded to, if only by indecision. No collective resolution by philosophers to ignore substantive moral issues will prevent such issues from being widely discussed, or reactive policies from being enacted. However, such a decision by the philosophers could well inhibit the clarity and intelligence of such discussions, or adversely affect the propriety and the efficacy of such policies.

In a recent address, Rawls urged moral philosophers to extend their interests beyond the traditional confines of their discipline, since:

The further advance of moral philosophy depends upon a deeper understanding of the structure of moral conceptions and of their connections with human sensibility We must not turn away from this task because much of it may appear to belong to psychology of social theory and not to philosophy. For the fact is that others are not prompted by philosophical inclination to pursue moral theory; yet this motivation is essential, for without it the inquiry has the wrong focus. (Oct., 1975, p. 22).

It is to Rawls's enduring credit that he has enriched contemporary moral deliberation by preparing an ambitious system of thought which combines some significant insights of traditional moral philosophy, the judgments of ordinary moral reflection, the findings of contemporary social and behavioral sciences, and the rigor of recent analytic philosophy. Furthermore, in justice as fairness, Rawls has devised a scheme of moral thought which, as a system, succeeds in avoiding the more obvious pitfalls of each of its components taken separately. First of all, while his system employs historical insights, it is far more than a simple recapitulation of archaic doctrines. Secondly, while Rawls's theory draws from the results of ordinary moral judgment, it does not reduce to mere relativism or conventionalism. Thirdly, like a good scientific theory, justice as fairness displays coherence, unity, and a

"vulnerability" to factual information. Moreover, mindful of recent developments in analytic philosophy, Rawls seeks conceptual clarity, and successfully avoids empty or metaphysically obscure locutions. And yet, while he uses conceptual analysis and scientific data to good effect, he does not, to borrow Wordsworth's phrase, "murder to dissect"; that is to say, the normative elements of his theory are neither dismantled by analytic reduction, nor are they absorbed, without a trace, into the structure of existing scientific disciplines.

Many reviewers and commentators have predicted that *A Theory of Justice* will prove to be a lasting contribution to moral philosophy. I agree. Even if Rawls's system fails to survive essentially intact, surely many elements thereof will endure. But more significantly, Rawls has been a prominent force in the recent move to bring normative moral philosophy -- the dedicated engagement of intellect in the search for human ends -- back into the contemporary scene. For this, as well as the insight, elegance, breadth, and rigor of his system, Rawls fully deserves the considerable attention and commendation that he has received in the brief four years since the publication of *A Theory of Justice*.

Thomas Nagel's eloquent tribute, with which I fully concur, bears repeating:

Rawls possesses a deep sense of the multiple connections between social institutions and individual psychology. Without illusion he describes a pluralistic social order that will call forth the support of free men and evoke what is best in them. To have made such a vision precise, alive, and convincing is a memorable achievement. (1973, p. 234).

NOTES

1. The reader interested in the critical responses to Rawls's theory may wish to examine Brian Barry's *The Liberal Theory of Justice* (1973), Norman Daniels' anthology, *Reading Rawls* (1975), and several of the many critical papers listed among the References at the end of this dissertation.
2. My personal view is that many modern philosophers who have dismissed ethical assertions as being subjective and cognitively meaningless expressions of emotion, inclination, or advice, have reached such conclusions after holding ethical claims up to criteria more appropriate to factual discourse. The predictable failure of moral discourse to meet these standards has led to the further, and I believe untenable, conclusion that moral assertions are "meaningless." In reply, I would urge that moral assertions, unlike empirical fact claims, are fundamentally *practical* and thus context dependent; that is to say, moral inquiry asks "What shall I do?", rather than "What is the case?" Furthermore, moral discourse is fundamentally *social*, rather than *individual*. Accordingly, the inclination of some philosophers to reduce meaning to personal data of immediate experience necessarily excludes the very dimensions of practical import and social context that give meaning to moral discourse. While I believe that Rawls would concur with much of this assessment, I do not believe that I have encountered this particular reply to the non-cognitivists in *A Theory of Justice*. Since this chapter is an exposition of Rawls's views, I will say no more here of my own position.
3. The term "rational contractors" was coined by R. M. Hare (1972).
4. Noam Chomsky's revolutionary new "transformational linguistics" rests a great deal of theoretical weight upon the remarkable grammatical sense of children. Chomsky sees this as evidence of an innate "universal grammar" that may be, so to speak, "pre-wired" in the human brain. However, this carries the analogy with moral theory too far. Rawls is not clearly proposing that there is a comparable, naturally "innate" morality.

Hampshire (1972) expresses the linguistic analogy with morality quite well: "The point of moral theory, and so of philosophical ethics, is to find some very general guiding principles that explain the apparently unconnected moral beliefs that constitute a prevailing morality. Rather as a linguist and philologist may look for the general principles that determine word order and the structure of sentences in English, so the philosophical moralist looks for the general principles, or the single principles, that explain the apparently diverse arrangements that we would consider unjust and therefore wrong. If we do succeed in finding such principles, which fit the facts of our ordinary moral beliefs fairly well except in a few marginal cases, then we can use the principles as a guide in doubtful cases" (p. 34).

(For Rawls's use of this analogy see p. 47 of *A Theory of Justice*.)

5. In his early paper, "Outline of a Decision Procedure in Ethics" (1951), Rawls presents a detailed list of criteria for "considered moral judgments." Other descriptions and criteria of effective moral deliberation are, of course, commonplace in the literature of moral philosophy. (for a statement quite harmonious with that of Rawls, see R. B. Brandt's, "Qualified Attitude Method," in his *Ethical Theory*, 1959, p. 250.)
6. This phrase "weak conditions" might be misleading. Rawls simply means that, other things equal, we should prefer the conditions in the original position to be general, uncomplicated, uncontroversial, and *few* -- "a constrained minimum" (p. 583). As Adina Schwartz puts it: "These conditions [of the original position] are strong enough to generate 'a workable theory of justice,' but they are weaker (in the sense of being more widely acceptable) than the conditions invoked in other theories' interpretation of the initial [contractual] situation" (1973, pp. 295-296).
7. Similar criticisms 'have come from Barry (1973), Lyons (1972), Margolis (1973), and Schwartz (1973), not to mention several others.
8. It is not my task to rescue Rawls from the critics of his general theory. However, I might suggest that a possible line of rebuttal to the critics of his claims of universality may be found in the works of his Harvard colleague, Lawrence Kohlberg. Kohlberg claims to have discovered culturally universal moral *structures* amidst an acknowledged diversity of moral *contents* (1967, p. 178). Might the principles of justice be held in reflective equilibrium with the moral judgments based upon these allegedly universal structures? Regretfully, haven't the space to pursue the question.
9. Several commentators suggest that Rawls's original position has a *justificatory* function as well as an "expository" function. Indeed, without this additional function, much of the point of this conceptual device is lost. Unfortunately, I must resist the temptation to enlarge this chapter with a consideration of this interesting question. For more discussion of this point see Carr (1975), Daniels (1975), Honderich (1975), Parekh (1972) and Schwartz (1973).
10. Rawls has expanded his views concerning the logical status of the "formal constraints of justice," in his presidential address to the Eastern Division of the American Philosophical Association: "The Independence of Moral Theory" (November 1975, pp. 11-14).
11. See especially the papers by Nagel (1973), Schwartz (1973), and Teitelman (1972). Rawls has replied to these critics on his recently published paper: "Fairness to Goodness" (October 1975, pp. 536-554).
12. Most of Rawls's section on "Fair Equality of Opportunity" (§14) consists of an insightful analysis of "pure procedural justice." This is still another of the many sections of the book which are illuminating in themselves, however much the whole system may or may not succeed. As with so many portions of this rich and comprehensive book, I must regretfully

pass over this topic.

13. I would guess that the difference principle has drawn more critical attention than almost any other aspect of Rawls's theory of justice. (The conditions of the original position and the priority of liberty have possibly attracted about as much response.) One of the most searching critiques is in the seventh chapter of Robert Nozick's book, *Anarchy, State and Utopia* (1974). (See especially p. 195.) Other papers dealing with this topic are too numerous for me to list them, and so I won't.
14. Russell made the remark in his Nobel Prize acceptance speech in 1951 (reprinted in Russell, 1962, p. 131).
15. "The Principle [of Efficiency, (also called "Pareto Optimality")] holds that 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. Thus a distribution of a stock of commodities among certain individuals is efficient if there exists no redistribution of these goods that improves the circumstances of at least one of these individuals without another being disadvantaged." (Rawls, 1971, p. 67).
16. Rawls has much more to say about just economic institutions (primarily in the first few sections of his fifth chapter). However, in the interest of space, I will have to exclude an account of this material.
17. When Rawls writes of a plan chosen with "full deliberative rationality," he means that "it is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires." (p. 417)
18. The reader may also perceive a parallel with Lawrence Kohlberg's theory of moral development. In *A Theory of Justice*, Rawls gives Kohlberg no more notice than two footnote citations (pp. 460n, 461-462n). Rawls acknowledges that his "morality of association" is parallel to Kohlberg's "stages" three to five. However, he feels that the superiority of the final "morality of principles" must be argued on philosophical grounds in that it cannot "be established on the basis of a psychological theory of development alone" (p. 462n). In contrast, Kohlberg is far more impressed with Rawls's theory of justice. For Kohlberg's views on the affinity of his theory with justice as fairness, see "The Claim to Moral Adequacy of a Highest State of Moral Judgment" (1973).
19. The debate continues. In an article just received, Stephen Darwall replies to Johnson in "A Defense of the Kantian Interpretation." (1976, pp. 164-170)
20. Quoted by Paul Samuelson in his column in *Newsweek*, Dec. 30, 1974, p. 54. I do not have the source in Hobhouse's works. I believe that this brief quotation from Hobhouse suggests the most fruitful approach toward a refutation of such radical libertarians as Ayn

Rand, John Hospers and Robert Nozick. It seems to me that these writers tend to treat an ongoing, functioning social system and a civilized condition as "free gifts" toward whose preservation and maintenance the individual owes virtually nothing. I personally find this position to be not only wildly misconceived but potentially dangerous.

21. Elsewhere, Rawls presents the issue this way: "The concepts of justice and goodness are linked with distinct principles and the question of congruence is whether these two families of criteria fit together. More precisely, each concept with its associated principles defines a point of view from which institutions, actions, and plans of life can be assessed. A sense of justice is an effective desire to apply and to act from the principles of justice and so from the point of view of justice. Thus what is to be established is that it is rational (as defined by the thin theory of the good) for those in a wellordered society to affirm their sense of justice as regulative of their plan of life. It remains to be shown that this disposition to take up and to be guided by the standpoint of justice accords with the individual's good." (p. 567)
22. The source in Nietzsche's work is: *Untimely Meditations: Third Essay: Schopenhauer as Educator*, sec. 6, cited from J. R. Hollingsdale, *Nietzsche: The Man and His Philosophy* (Baton Rouge, Louisiana State University Press, 1965), p. 127.
23. To be sure, not all critical philosophers would agree. For an excellent indication of the clash between Rawls's normative ethics and the critical-analytic approach to moral philosophy, see R. M. Hare's two-part review of Rawls's book (Hare, April 1973, July 1973), and Spencer Carr's reply in Rawls's behalf (1975).