RAWLS' THEORY OF JUSTICE

THOMAS M. SCANLON, JR.†

Rawls' book is a comprehensive and systematic presentation of a particular ideal of social life. The aim of the book is to analyze this ideal in a way that allows us to see clearly how it differs from prominent alternatives and on what grounds it may be preferred to them. In carrying out this analysis Rawls presents and draws upon not only a theory of distributive justice and a theory of political rights, but also a theory of value, a theory of obligation and a theory of moral psychology.

Contemporary political philosophy has already been altered by Rawls' book, but the leading ideas of his theory are in a number of respects familiar ones. They are familiar within the philosophical community since they have been set forth by Rawls in a series of important and influential articles over the course of the last fifteen years.[1] Rawls has altered and clarified important points in his argument since these articles first appeared, and the book presents a large amount of new material, but the main thrust of Rawls' theory can be seen in these shorter works. The idea of Rawls' book will also seem familiar to many who have not read his articles, for the ideal of social life he describes is one that lies at the heart of liberal political theory, and the principles and policies which this ideal supports are, in general, ones that liberals have traditionally supported. Finally, the central analytical device in Rawls' argument is a variant of the familiar idea of a social contract. Such a contract is perhaps most often thought of in connection with accounts of the origins of political authority. While Rawls does give an account of political obligation (one which does not require actual consent) his use of the idea of accepting institutions in an initial contractual situation is not confined to this purpose but serves more broadly as the basis for critical appraisal of economic,

† Associate Professor of Philosophy, Princeton University. A.B. Princeton University, 1962; Ph.D. Harvard University, 1968.

[1] Rawls' basic thesis about justice was set forth in *Justice as Fairness*, 67 PHIL. REV. 164 (1958). This article has been widely reprinted, *e.g.* in PHILOSOPHY, POLITICS AND SOCIETY: SECOND SERIES 132 (P. Laslett & W. Runciman eds. 1962) and in JUSTICE AND SOCIAL POLICY 80 (F. Olafson ed. 1961). His ideas on liberty were developed in *Constitutional Liberty and the Concept of Justice*, in NOMOS VI: JUSTICE 98 (C. Friedrich & J. Chapman eds. 1963). *Distributive Justice*, in PHILOSOPHY, POLITICS AND SOCIETY: THIRD SERIES 58 (P. Laslett & W. Runciman eds. 1967) outlines economic institutions which would satisfy Rawls' conception of justice; *The Justification of Civil Disobedience*, in CIVIL DISOBEDIENCE 240 (H. Bedau ed. 1969) presents his theory of political obligation. Finally, Rawls' account of the development of moral sentiments is set out in *The Sense of Justice*, 72 PHIL. REV. 281 (1963).

social and political institutions. Rawls' work thus develops and carries forward in a systematic way that branch of contract theory represented by Kant and Rousseau, who saw the idea of a hypothetical initial agreement as a necessary condition for the legitimacy of political institutions.

In this study I will present and discuss what I take to be the most important arguments of Rawls' book, emphasizing in each case the way in which these arguments are related to the ideal of social life which forms the core of his theory. I will begin with a discussion of Rawls' method and overall aims. In Section II I will consider his arguments against what he calls perfectionism and in favor of his thesis of the priority of liberty. Section III is concerned with his arguments against utilitarianism and Section IV with his thesis that social and economic inequalities are just only if they work to the advantage of the worst-off members of society.

I. RAWLS' PROGRAM AND METHOD

Principles of justice, according to Rawls, are principles which "provide a way of assigning rights and duties in the basic institutions of society and . . . define the appropriate distribution of the benefits and burdens of social cooperation."[2] One can think of principles of justice as instruments of theoretical criticism and practical choice, guiding our appraisal of different social institutions, but it is essential to think of such principles also as one of the crucial operative elements in a functioning set of institutions. This is the perspective taken by Rawls and expressed in his notion of a well-ordered society.

> Now let us say that a society is well-ordered when it is not only designed to advance the good of its members but when it is also effectively regulated by a public conception of justice. That is, it is a society in which (1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles. In this case while men may put forth excessive demands on one another, they nevertheless acknowledge a common point of view from which their claims may be adjudicated. If men's inclination to self-interest makes their vigilance against one another necessary, their public sense of justice makes their secure association together possible. Among individuals with disparate aims and purposes a shared conception of justice establishes the bonds of civic friendship; the general desire for justice limits the pursuit of other ends. One may

[2] J. RAWLS, A THEORY OF JUSTICE 4 (1971) [hereinafter cited as RAWLS].

> think of a public conception of justice as constituting the fundamental charter of a well-ordered human association.[3]

The idea of an agreement on principles or procedures which stands behind conflicts of interest and provides the basis for their resolution is, in various forms, a familiar one. One may think of such an agreement as something which just grows up in fortunate circumstances or as something inherited from tradition, but the content of such an agreement is clearly subject to rational scrutiny. It is a comprehensible question whether some principles of justice are rationally to be preferred to others for the crucial social role Rawls assigns to them. This question is not immediately answerable, however, without some consideration of what the appropriate standards are for such a far-reaching and headily abstract choice.

Rawls' approach to this problem is analogous to what must have been the dream of many an arbitrator: to separate the contending parties in a dispute and to force them somehow to come to an agreement on principles which they would accept in general as a basis for the adjudication of disputes of the kind in which they are involved. In arranging such an argument-within-an-argument an arbitrator might have a number of different aims in view. First, as a practical matter, he wants the parties to come to some agreement on principles and he wants this agreement to be one that will actually hold up when the principles are applied to particular cases. In addition, he may have his own views as to which principles should be chosen, views about particular substantive decisions these principles should or should not dictate, and views about the kinds of considerations which it is relevant for such principles to give weight to and the kinds of considerations good principles should ignore. The arbitrator's problem is to devise the ground rules for the argument-within-an-argument and provide incentives to the parties in such a way as to ensure that these aims will be met.

What Rawls calls the Original Position can be seen as a construction on this model, and the chief strategic device it employs to achieve its aims is the notion of choice behind a partial veil of ignorance. The parties in the Original Position know that they are contemporaries in some society. While they do not know any details of the circumstances prevailing in this society, they do know that these circumstances are ones in which cooperation is both possible and necessary, and they have full access to general truths of social science telling them how social institutions work and the difficulties to which they are subject.

---

[3] *Id.* 4-5.

The parties do not know their own places in their society, their particular talents and skills or even their own conceptions of the good (*i.e.* their tastes, goals and objectives). They are supposed to be rational and to be mutually disinterested in the sense that in their choice of principles they are not motivated either by sympathy or by envy, but rather each by a desire to do as well for himself as he can.

Since the parties do not know their own conceptions of the good, there is a problem of specifying the terms in which they are to understand judgments of relative well-being. Rawls' strategy here is to focus on certain categories of goods which it is rational to want no matter what one's conception of the good may be. A good may have this property either because, like self-respect, it has a central place in any conception of the good or because, like income and liberty, it is necessary as a means to most other things one may desire. Some goods which are neutrally desirable in this sense, such as a sound constitution or good eyes, are not things whose allotment is a function of social institutions (although they are to some extent dependent on the circumstances of one's life). Neutrally desirable goods whose allotment is directly dependent on social institutions are what Rawls calls primary social goods.[4] He lists these as rights and liberties, powers and opportunities, income and wealth, and the bases of self-respect. Since Rawls uses primary social goods as an index of relative well-being, he needs to claim not only that these goods are things which it is rational for anyone to want no matter what his conception of the good may be, but also that there is at least a rough ranking of "bundles" of such goods which is similarly neutral, *i.e.*, a ranking of combinations of primary goods (so much income, certain liberties, etc.) as "greater" and "smaller" such that it is rational for anyone to prefer a "greater" allotment of primary social goods to a "smaller" one.[5]

In addition to wanting for themselves the largest possible bundle of primary social goods, the parties are assumed to be motivated to choose principles which it will be rational and psychologically possible for them to adhere to, and principles which will provide the basis for a stable and lasting social order by generating their own psychological base of support; *i.e.*, it is supposed that the parties will make their choices only after they have determined that persons growing up in a society governed by their chosen principles will naturally acquire a sense of justice motivating them to act in accord with those principles.

---

[4] *Id.* 62, 92.

[5] This problem, which Rawls calls the index problem for primary goods, is discussed at *id.* 93-95.

Rawls puts forward the following Two Principles as the principles of justice that would be chosen in the Original Position:

> [*First Principle*]
> Each person is to have an equal right to the most extensive total system of basic liberties compatible with a similar system of liberty for all.[6]
>
> [*Second Principle*]
> Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity.[7]

It would be a mistake to see Rawls' argument as proceeding deductively, first from the description of the role of a conception of justice in a well-ordered society to the idea of the Original Position as a means for choosing the (instrumentally) best such conception, and thence to the Two Principles of justice as the particular conception which passes this test. To construe Rawls' argument in this way would be a mistake first because, as he emphasizes,[8] his argument does not proceed *only* in this direction. The choice of the constraints which define the Original Position is guided initially by the idea of the role that justice is to play in a well-ordered society and by strategic considerations designed to insure that some agreement is reached. But these constraints are also trimmed and shaped to insure that the Original Position will yield results which conform with our considered judgments of what is just and unjust in particular cases. What is sought is not a *proof* of the Two Principles but a setting out of their relations, on the one hand, to our considered judgments of justice and, on the other, to certain general ideas of social cooperation. Each of the elements in such a total picture, and the way in which all of them fit together, elucidates and provides support for the others. Such a method of reasoning is not viciously circular since it is not assumed, nor is it by any means obvious from the outset, that our considered judgments of justice and our most general notions about social cooperation can be fitted together in a systematic and cohesive way.

The second reason why it would be a mistake to see Rawls' strategy as deductive is that some of the most likely alternatives to the

---

[6] *Id.* 250.

[7] *Id.* 83. This principle is advanced as the favored interpretation of the more ambiguous principle that "social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all." *Id.* 60. On the relation of these two formulations of clause (a), *see* note 68 *infra.*

[8] *Id.* 20, 21.

Two Principles of justice appear to be ruled out by the form of the Original Position itself or by the very idea of a well-ordered society as one in which conflicts are regulated by a notion of justice in the way Rawls describes. Two examples will be helpful here: first, it may seem to many that to take as an ideal a "well-ordered society" in which conflicts of interest arise and are settled with reference to principles of justice chosen on the basis of their appeal to parties whose main motivation is each to secure as much as possible for himself is already to build in too many features of the societies with which we are most familiar. Ruled out without serious consideration, it may be thought, is the possibility of a society in which the sources of major conflict are largely eliminated and in which the basis of association lies in relations of love and sympathy rather than in principles of distributive justice.

Second, the possibility of a society in which the ruling principles are given by a particular religious ideal or a particular secular ideal of human excellence appears to be ruled out in the Original Position by the requirement that the parties are ignorant of their own conceptions of the good. But if judgments of relative value are not just matters of taste and opinion but also of objective fact, then it appears to be irrational (and tendentious) to block consideration of such facts from the deliberations of the parties in the Original Position.

It would not be unreasonable to say that these two objections represent alternative social ideals—which I will call, respectively, communitarian and perfectionist—which are not merely rivals to Rawls' Two Principles of justice for consideration in the Original Position but rivals to the outlook represented by the Original Position itself. If Rawls is not to be construed as begging the question against these alternatives, then he cannot be construed as following the deductive strategy outlined above.

Rawls' response to the communitarian challenge proceeds along two related lines. First he argues that it is a mistake to see love and sympathy as in themselves rivals to justice as guides for social action. These feelings may move us to sacrifice for others, but in themselves they provide no guide to the degree of sacrifice appropriate, no guide as to which of two loving persons should in a given case make way for the other, and, most clearly, no guide to how one should act when the interests of two loved persons come into conflict.[9] Second, Rawls argues that the natural attitudes of love and trust themselves presuppose a conception of what it is to respect another as a distinct person, *i.e.* a

[9] *Id.* 191.

notion of justice of the kind which the Original Position construction is designed to capture.[10] If these arguments are correct, then a communitarian ideal of social life of the kind I have described is not, at the fundamental level, an *alternative* to cooperation on terms of justice. It consists rather in putting new flesh on the bones which a theory of justice provides.

Differences between the two outlooks remain, however, which are too deep to be called mere matters of emphasis. The communitarian may be understood as asking in part how far a healthy society could rely upon justice in its distributive sense as the main counterforce to certain tensions and instabilities. The most important question, the communitarian contends, is not whether the institutions maintain strict distributive justice; it is whether they provide a basis for healthy human relations or whether, on the contrary, they foster social relations which are antithetical to the growth of natural attitudes of love, sympathy and trust. Rawls argues that the connection between justice and the values of community is much more intimate than this criticism would suggest. Through the adoption of his Two Principles as a public conception of justice, Rawls argues, the members of a society express their respect for one another as moral persons; moreover, since the Second Principle constitutes "an undertaking to regard the distribution of natural abilities as a collective asset so that the more fortunate are to benefit only in ways that help those who have lost out,"[11] this principle embodies an idea of fraternity.[12] Thus, he maintains, a well-ordered society founded on these principles is a community in a strong sense, a social union in which each person may pursue his own good within a form of association which is itself a good for all.[13] I cannot consider these arguments in detail here; some will be treated at greater length below in the discussion of the Two Principles themselves. My present purpose in this example has been merely to indicate how the force of Rawls' theory depends not only on the details of the derivation of his principles in the Original Position, but also and more importantly on the coherent view of social cooperation which the theory as a whole provides, in this case particularly on the related accounts of moral psychology, of the concept of a person and of the idea of the good.

---

[10] *Id.* 486-87.
[11] *Id.* 179.
[12] *Id.* 105.
[13] *Id.* 522-29.

II. Liberty

A. *The Argument Against Perfectionism*

Those theories which Rawls calls perfectionist direct us "to arrange institutions and to define the duties and obligations of individuals so as to maximize the achievement of human excellence in art, science, and culture."[14] In my previous remarks I have grouped such theories together with theories which take as the ruling aim of social institutions the promotion of a particular religious ideal. This grouping may seem somewhat unfair since there is in theories of the first sort a strong tendency toward elitism—*i.e.* towards placing much greater emphasis on the needs and interests of some members of society than on those of others—and while some religious-based theories may exhibit a tendency of this kind in singling out a small group (*e.g.* "the elect" or the clergy) for special privileges, this need not be regarded as a characteristic feature of the type.

What all of these theories, religious and secular, share is first of all a teleological structure:[15] once the value of a certain end is established, social institutions are to be appraised strictly on the basis of their tendency to promote this end. In addition, quite apart from tendencies to elitism, all of these theories raise serious problems concerning individual liberty: institutions which preserve the opportunity for each person to adopt and pursue his own interests and ideals and to try to persuade others to follow him will be justified on perfectionist grounds only to the extent that they are the most effective means to the promotion of the given end.

Now it would be possible to reject theories of this kind simply on the basis of their tendency to support institutions which conflict with our considered judgments of justice, and then to design the Original Position in such a way that the offending theories are ruled out. Adopted alone, however, this strategy is not wholly satisfying. If we can give no independent rationale for the design of the Original Position then this maneuver appears somewhat ad hoc. To provide such a rationale, based on a non-perfectionist ideal of social cooperation, would not constitute a refutation of perfectionism; but without such a rationale we are left with no response to the basic theoretical challenge which these theories raise: If there is an objective difference in the intrinsic value of different talents, goals and pursuits why should not information about these differences be used by the parties in the Original Position as the basis for their choice of the principles by which social

[14] *Id.* 325.

[15] The notion of a teleological theory is discussed more fully in Section IV *infra*.

institutions will be judged? How, in short, can we defend an egalitarian or libertarian position without embracing some form of skepticism about values?

Rawls' response to this challenge (and his rationale for the design of the Original Position) is grounded in the notion that social institutions are just only if they can be defended to each of their members on the basis of the contribution they make to his good as assessed from his point of view. We must be able to say to each member that the arrangements he is asked to accept provide as well for him as they possibly can, consistent with satisfying the parallel demands of others. In order to spell out this idea more fully it is necessary first to consider Rawls' analysis of the notion of an individual person's good.[16]

Those experiences, ends and activities are components in the good for a particular individual, Rawls argues, which have an important place in a plan of life which it would be rational for him to choose. Now it may seem that a person could be said rationally to choose a plan of life (if at all) only after he has developed a conception of his own good, on the basis of which he can judge and rank alternative plans of life. But Rawls argues, persuasively I think, that this is not the case. In real life our deliberations about those actual choices which, taken together, determine our plan of life proceed on the basis of knowledge of our present tastes and capacities, knowledge of what things we have in the past found satisfying, and knowledge of general principles governing the ways in which our tastes and capacities are subject to growth and change over time. This information allows us to decide on courses of action not only with the aim of satisfying our current desires but also with the knowledge and intent that our choices will be instrumental in determining what interests, talents and desires we will come to have in the future. Long range choices such as the choice of a career or a place to live give perhaps the best example of choices which, because they may be foreseen to have far-reaching effects on our interests and objectives, must be made on some basis which goes beyond the satisfaction of our current desires and specific interests.

Rawls puts forward a negative and a positive thesis about this process of deliberation. The negative thesis consists of an attack on the idea that there must be some single overriding general goal (*e.g.* the maximization of satisfaction or happiness) which underlies all of our deliberations and explains how we can compare and choose between disparate alternatives.[17] The positive thesis consists of a sketch

[16] *See generally* RAWLS 60-65.
[17] *Id.* §§ 83-84.

of standards of rationality with reference to which our choices, particularly those most general and far-reaching choices described as choices between alternative life plans, can be criticized. This sketch consists of two parts. First, there are general principles of rational choice according to which it is irrational, *e.g.*, for anyone to prefer plan of life *A* to plan of life *B* if *B* involves the development of exactly the same interests and desires as *A* and provides for their satisfaction at a markedly higher level. Not all of these principles, which Rawls calls "counting principles," are as uncontroversial as this example, but all are fairly weak, and taken together they by no means can be expected to determine a unique plan as the only rational choice for a person to make. A choice from among the plans not ruled out by these principles (the set of maximal plans) will involve such things as comparing the relative intensity of different desires and the relative value for us of different kinds of accomplishments. For this choice there are on Rawls' view no principles of rationality which directly require a choice of some plans over others. The only relevant standards concern the manner in which the choice is made—whether the relevant evidence has been duly weighed, the possible sources of uncertainty and error properly discounted for, etc. These criteria are grouped together by Rawls under the heading "deliberative rationality."

Thus, to say that a certain thing is, objectively, a good for a certain person is, on Rawls' analysis,[18] to say that it would be a prominent feature in a plan of life which that person would hypothetically choose, with deliberative rationality, from among the class of maximal plans. Under any actual conditions, of course, not only the means for attaining those things which are goods for us, but also ideal conditions for determining what things are such goods, will in some measure be lacking. What the parties in Rawls' Original Position look for in a society is not only the means for securing those things, whatever they may be, which are objectively components of their good, but also the conditions necessary for determining what these goods are.

From the fact that the parties in Rawls' Original Position suppose that as members of a society they will choose their own plan of life, and hence also determine their own conception of the good, it should not be thought that they suppose themselves to be independent of social forces which will in large part shape and influence the choices they make. It would be idle to deny that such influences exist, and irrational to object to all such influences as interfering with one's liberty. But it is still reasonable to prefer some institutions to others

[18] *Id.* 417.

on grounds of the conditions they provide for rationally forming a conception of one's good. Obviously one may reasonably object, simply on grounds of efficiency, to institutions which place arbitrary obstacles and difficulties in the way of individuals' attempts to get a clear view of the alternatives open to them, of their own potentialities, and of what they and others can expect from various courses of action. A more difficult case is presented by the fact that some features of institutions will not merely be random inferences but can be seen clearly to favor certain choices and to discourage others, and to do this not by just enlarging people's views or by approaching "ideal conditions" thereby favoring "the correct answer," but rather by skewing the evidence available or by restricting the alternatives likely to be considered, or by affecting people's deliberations in other more subtle and indirect ways. Systematic interference of this kind might be the result of relatively fixed impersonal features of institutional arrangements. Alternatively, certain individuals may be charged with overseeing and maintaining these influences through censorship or other devices.

It is one of the features of perfectionist views which strike us intuitively as objectionable that such views may authorize the use of means of this sort in order to produce individuals conforming to a particular ideal. Now we cannot simply reject as involving unacceptable "conditioning" all social institutions which mold a person's choices and beliefs without his consent with the aim of bringing him closer to some ideal. Certainly Rawls cannot do this. For as he himself says, his own view involves a certain ideal of the person, and he is at some pains to show[19] that there are psychological laws which give us reason to believe that persons growing up in a well-ordered society governed by his Two Principles of justice will naturally acquire what he calls a sense of justice—the tendency to understand and be motivated by considerations of justice as specified by those principles. The action of these psychological laws is in part dependent upon the intellectual activity of the person on whom they are acting, but is also in large part something which happens to a person without his knowledge or rational scrutiny.

How then is one to distinguish among the various ways in which social institutions may be arranged to influence the choices and beliefs of their members without each member's consent? Can one distinguish acceptable from unacceptable influences of this kind on any basis other than an appraisal of the relative value of the particular types of persons these influences produce? The appropriate standards for making

[19] *Id.* §§ 51-59.

this distinction on Rawls' theory seem to me to be suggested by the criteria he offers for distinguishing justifiable from unjustifiable paternalism.[20] The relevant principles here require first that paternalistic interventions, *i.e.* interventions in a person's life "for his own sake" which are pursued contrary to his wishes or without his knowledge, have to be rationally justifiable *to him* after the fact. Second, such interventions must be justified on the grounds that the subject's evident failure or absence of reason and will at the time rules out a direct presentation of the issues to him for his own rational consideration and decision. A third requirement is that the intervention "must be guided by the principles of justice and what is known about the subject's more permanent aims and preferences"[21] or, failing such knowledge, by some neutral standard such as that provided by the primary goods.

While Rawls formulates these requirements specifically for the case of paternalistic action by one person toward another, they seem to be applicable as well to the broader class of interventions we are considering. This is indicated, for example, in the fact that Rawls' defense of the process by which a sense of justice is inculcated in persons who grow up in a well-ordered society governed by his Two Principles of justice advances considerations essentially parallel to these requirements.[22] One can maintain here, first, that the principles which form the content of this sense of justice are ones the person can later come to see as justified. (This fact alone, of course, would not be an adequate defense since any successful piece of indoctrination, or at least any successful indoctrination of justifiable beliefs, could make this claim.) Further, the practices of moral education in a well-ordered society proceed as far as possible by appeal to the subject's reason, and rely upon other factors only insofar as the natural limitations of childhood make necessary. Finally, the acquisition of a sense of justice is, it is argued, not inconsistent with a person's good. Since the conception of justice (*i.e.* Rawls') which is the content of the sense of justice in question provides a secure protection for each person's interests and for his desire to determine his own conception of the good, the acquisition of such a sense of justice is not something which leaves a person open to exploitation or manipulation by others. In addition, having a sense of justice is a necessary condition for sharing fully in the life of a well-ordered society[23] and a necessary condi-

[20] *Id.* 249-50.
[21] *Id.* 250.
[22] *Id.* 514-15.
[23] *Id.* 571.

tion as well for susceptibility to the natural attitudes of friendship, love and trust.[24] These are things, Rawls argues, which almost[25] anyone has reason to want.

Without going fully into the arguments for these claims, we may compare them to the case which might be made on perfectionist grounds for features of social institutions designed to mold or restrict the choices of their members so as to promote a particular secular or religious ideal. There is a clear sense in which such features will have a rational justification: they will be justifiable on the basis of the objective value of the particular ideal in question. A perfectionist might thus maintain that the interferences with a person's liberty which these features represent are ones which he should, rationally, come to accept. But the justification which is offered by the perfectionist will not necessarily be one which claims that these features promote the good of the person whose liberty is restricted or which claims that they are consistent with his desire to determine his own conception of the good; it is apt to appeal instead to some impersonal scheme of values. Moreover, this justification need not be based on considerations which would be agreed upon by almost anyone regardless of his conception of the good. Rather, it is likely to be based on one specific conception of the good which, even if it is objectively correct, may nonetheless be something which is a matter of some disagreement among rational adults in the society in question. Indeed, it is just the fact that this conception of the good, though correct, does not compel general agreement, which may be taken on perfectionist grounds to make necessary the intervention in question. On Rawls' theory, however, such interventions are permitted only when there is "evident failure or absence of reason and will," a phrase intended to cover cases such as infancy, insanity or coma which involve major diminution of rational capacities relative to the standard of "a normal adult in full possession of his faculties."

Thus, while Rawls' theory bases principles of justice on a hypothetical choice made by persons who may appear to be standing temporarily outside any particular society, the point of view which the theory takes as fundamental is actually that of a person *in* society. The parties in the Original Position do not act from special wisdom or knowledge which enables them to make choices which they later, as persons under the limiting and distorting conditions of real life in an actual society, will have to take on faith. Rather, the parties' aim

[24] *Id.* 570.

[25] Rawls does allow for the possibility that there may be "some persons for whom the affirmation of their sense of justice is not a good." *Id.* 575.

is to make choices which they, as real citizens, will have reason to accept. Each party therefore regards his own judgment as a real citizen as sovereign—not as infallible or immune from limitations, but as the basis from which his life will be lived, his choices made and his work as ideal contractor appraised.

Rawls remarks that "embedded in the principles of justice there is an ideal of the person that provides an Archimedean Point for judging the basic structure of society."[26] Although I have not described this ideal in full, the preceding argument seems to me to illustrate part of the force of this remark. The ideal of each person as a rational chooser of his own ends and plans provides an Archimedean Point partly in virtue of the fact that this conception of a person is taken to be prior to any particular independently-determined conception of his good. One need not be a skeptic about values or truth to hold that each of us does in fact look at himself in this way. If this is so, then the assumption that the parties in the Original Position adopt this view of themselves should seem a natural one, and the fact that certain principles of social cooperation involve the recognition of each member of society as in this sense a sovereign equal, while others involve the denial of this status to at least some members, should seem a fact of some importance.

The conception of the person described by Rawls is of course not an Archimedean Point in the sense of being itself a notion formed outside of or independent of particular social and historical circumstances. It may well be that this conception of the person and the ideal of social cooperation founded on it are typical of particular historical eras and civilizations. But this is not in itself an objection to Rawls' theory, particularly if, as it seems to me, the conception of the person in question is one that has a particularly deep hold on us and is not a matter of great controversy or of significant variation across the range of societies to which the theory should be expected to apply. The question is not whether this conception of the person is in some sense absolute, but whether the particular features of this conception that are appealed to in Rawls' argument are more controversial than the conclusions they are used to support.

Certainly this conception of the person involves a number of important parameters which must be fixed before the notion can be appealed to in support of conclusions about the justice or injustice of particular institutions. The most obvious of these is the standard of rationality: what is to count, for example, as "evident failure or

[26] *Id.* 261-65, 584.

absence or reason or will"? Other parameters are represented by the general facts of social science which the parties in the Original Position use in reaching their conclusions, by the notion of the primary social goods, and by other appeals to the idea that certain goods or circumstances are to be desired "no matter what one's conception of the good may be." The latter appeals depend upon some idea of the normal range of variation in conceptions of the good and upon some idea of the means and conditions required for the pursuit of these goods. All of these may be subject to some variation over time and social circumstances. But here again the theory need make no claim to *absoluteness* in these matters. It is sufficient to ask whether the appeals the theory makes to facts about persons and the circumstances of human life are controversial *for us*; in particular, whether the facts appealed to are more controversial than the conclusions at issue; and finally, whether the ways in which conclusions about the justice of institutions are made to depend on such facts strike us as plausible.

Much of the preceding discussion has been internal to Rawls' particular conception of social cooperation and is thus not in any proper sense a *refutation* of perfectionism. It is, rather, a description of an alternative ideal of social life, one which might be called "cooperation on a footing of justice." The development of this ideal enables Rawls to move beyond the observation that perfectionism seems to support arrangements which are at variance with our intuitive judgments of justice to a theory which explains why this should be so and provides a point of view from which we can see how the perfectionist challenge can be answered.

B. *The Priority of Liberty*

I turn now from these theoretical issues to consideration of Rawls' more specific conclusions concerning the place of liberty in just institutions. Rawls' substantive account of justice is put forward in two forms which he calls respectively the General Conception and the Special Conception of Justice as Fairness. The General Conception of Justice as Fairness provides that "[a]ll 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."[27] The Special Conception is expressed in the two principles of justice stated earlier,[28] with the proviso that the First Principle is to be held prior to the

[27] *Id.* 303.

[28] Text accompanying note 7 *supra.*

Second in a sense to be discussed more fully below. The Second Principle allows for inequalities in the distribution of goods other than basic liberties on terms similar to those specified by the General Conception, but the First Principle lays down a more stringent requirement of equality in basic liberties, a requirement which is not to be set aside for the sake of greater economic or social benefits. This principle and the rule specifying its priority receive their final statement in the following form.

> *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.
>
> *Priority Rule*
>
> The principles of justice are to be ranked in lexical order[29] 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, and (b) a less than equal liberty must be acceptable to those citizens with the lesser liberty.[30]

The "basic liberties" with which the First Principle is concerned are specified by Rawls as follows.

> The basic liberties of citizens are, roughly speaking, political liberty (the right to vote and to be eligible for public office) together with freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person along with the right to hold (personal) property; and freedom from arbitrary arrest and seizure as defined by the concept of the rule of law. These liberties are all required to be equal by the first principle, since citizens of a just society are to have the same basic rights.[31]

A liberty in the sense in which Rawls uses the term is defined by a complex of rights along with correlative duties of others to aid or not to interfere. Thus, by a restriction on liberty or an unequal liberty he means a restriction or inequality in what people are legally entitled to do (or, perhaps, entitled to do by the nonlegal rules defining the basic institutions of their society). Inequalities in people's ability

---

[29] "This is an order which requires us to satisfy the first principle in the ordering before we can move on to the second . . . . A principle does not come into play until those previous to it are either fully met or do not apply." RAWLS 43. For Rawls' initial statement of the Second Principle, see text accompanying note 7 *supra*.

[30] RAWLS 250 (footnote added).

[31] *Id.* 61.

to take advantage of their rights due, *e.g.*, to unequal economic means do not count as inequalities in liberty for Rawls but rather as inequalities in what he calls the "worth" or "value" of liberty. While the basic liberties must be held equally, the worth of these liberties may vary since any significant inequality in wealth, income or authority (allowed under the Second Principle) will represent an inequality in the ability of citizens to make use of their liberty in order to advance their ends.[32] Rawls stresses at a number of points[33] the importance of preserving "the fair value" of the basic liberties, particularly political liberties, but strict equality in the worth of these liberties is not required by the First Principle itself.

Two examples, frequently cited by Rawls, of restrictions on basic liberties that are justified on the ground that they strengthen the total system of basic liberty are the restrictions on the scope of majority rule imposed by a bill of rights and the restrictions on the freedom to speak imposed by a system of rules of order. In the first case a restriction of the legal powers of citizens is justified by the fact that more extensive powers could legally erase other basic liberties. In the second case what are sometimes called restrictions as to time, place and manner are imposed on the exercise of a basic liberty in order, Rawls says, to preserve the worth of that liberty to all.[34] Thus it appears that while equal worth of the basic liberties is not required by the First Principle, securing the worth of these liberties is one of the goals which can justify restrictions on basic liberties under the Priority Rule.[35]

1. The Preference for Basic Liberties over Other Primary Goods

Given the degree to which the content of the Priority Rule, and hence the claim of Rawls' theory to provide a secure basis for liberty, depends upon the distinction between the basic liberties and other goods and opportunities, it may seem surprising that no theoretical account of this distinction is offered. The list of familiar constitutional categories given above is offered by Rawls not as a precise enumeration of the class of basic liberties but only as indicating "roughly

---

[32] *Id.* 204.

[33] *E.g.*, *id.* 224-26, 277-78.

[34] *Id.* 203.

[35] Both of the cases considered here are examples of a lesser but still equal liberty as provided for by clause (a) of the Priority Rule. Primary examples of unequal liberty allowed by clause (b) seem to be cases of "justifiable paternalism" in which a less than equal liberty is "acceptable to those whose liberty is restricted" in the sense spelled out in the requirements discussed above. *See* text accompanying notes 19-20 *supra*.

speaking" what this class is to include. I suspect that here, as with the class of primary goods itself, no precise theoretical demarcation can be given. What is claimed for these liberties is just that, due both to the importance for anyone of the interests they safeguard and to their great instrumental value for the enjoyment of other goods, they are not only things it is rational for anyone to want but also things it is rational for anyone to value particularly highly relative to other primary social goods.

It is not claimed that these liberties are always to be valued more highly than any other goods. Rawls allows that under particularly dire conditions, when bare survival or the pursuit of the means for a minimally comfortable life is the dominant concern, and when the necessary prerequisites for the effective exercise of the basic liberties are lacking, it may be rational to sacrifice basic liberties for the sake of other goods such as increased security or economic development. It is under such conditions that the General Conception of Justice as Fairness applies. Rawls argues,[36] however, that as conditions improve and the possibility for the effective exercise of the basic liberties becomes real, people will set an increasingly high marginal value on basic liberties relative to other goods. After the most urgent wants are satisfied, people come to set greater importance on the liberty to determine and pursue their own plans of life. They will therefore insist on the right to pursue their own spiritual and cultural interests, seek to "secure the free internal life of the various communities of interests in which persons and groups seek to achieve . . . the ends and excellences to which they are drawn" and, in addition, "come to aspire to some control over the laws and rules that regulate their association, either by directly taking part themselves in its affairs or indirectly through representatives with whom they are affiliated by ties of culture and social situation."[37] Recognizing these tendencies, the parties in the Original Position will see that "[b]eyond some point it becomes and then remains irrational [for them] . . . to acknowledge a lesser liberty for the sake of greater material means and amenities. . . ."[38] Thus the position of liberty under the Special Conception makes explicit the priority that emerges under the General Conception as the natural preference for basic liberties over increases in other primary social goods asserts itself.

There are a number of questions one might raise concerning this argument. First, since the appeal to an increasing preference for

[36] RAWLS § 82.
[37] *Id.* 543.
[38] *Id.* 542.

basic liberties over other primary social goods represents Rawls' most detailed claim about the way in which the parties in the Original Position would order bundles of primary social goods, it naturally gives rise to questions of the sort considered above under the heading of "parameters." Rather than to consider the general question of whether this preference is in some suitable sense "universal," however, it seems to me more profitable to ask whether an appeal to such a preference provides adequate and interesting answers to those questions about liberty (and about the particular basic liberties listed by Rawls) that one would want a philosophical theory of liberty to answer.

Foremost among these is the question to what extent the basic liberties have some kind of absolute status and to what extent, and within what limits, they are to be understood and interpreted in terms of a balancing of competing interests. Rawls appears to have two answers to this question. The first, given by the Priority Rule, makes the limitation on acceptable balancing depend upon the distinction between basic liberties and other primary social goods: basic liberties are to be limited only for the sake of the total system of basic liberty itself. The second answer, and the one most often used by Rawls to indicate when a lesser but still equal liberty is just, is given by what he calls the Principle of the Common Interest:

> According to this principle institutions are ranked by how effectively they guarantee the conditions necessary for all equally to further their aims, or by how efficiently they advance shared ends that will similarly benefit everyone. Thus reasonable regulations to maintain public order and security, or efficient measures for public health and safety, promote the common interest in this sense. So do collective efforts for national defense in a just war.[39]

Rawls does not formulate this principle explicitly, but his discussion[40] suggests the following formulation: basic liberties may be restricted only when methods of reasoning acceptable to all make it clear that unrestricted liberties will lead to consequences generally agreed to be harmful for all.

Rawls seems to hold[41] that these two doctrines are consistent, *i.e.* that cases in which a restriction of basic liberties is justified by the Principle of the Common Interest are also cases in which basic liberty is being limited for the sake of the total system of basic liberty itself.

---

[39] *Id.* 97.

[40] *Id.* 213-14.

[41] *Id.* 246-47, 212-13.

This appears to be true in the most apocalyptic cases, *e.g.* cases in which a restriction of basic liberties is necessary as part of the common defense against an invasion. It may be true as well in some more mundane cases, such as Rawls' example of the restrictions imposed upon the right to speak by fair rules of order (taking into account, as was noted above, that what is protected in this case is not, strictly speaking, liberty but rather the worth of liberty). But if the restrictions on utterances imposed by such a set of rules count as restrictions on a basic liberty, then so also must similar restrictions on the time, place and manner of political demonstrations, religious festivals, parades, the placing of posters and the use of loudspeakers and sound trucks. Regulation of these activities is normally thought to be acceptable, and appears to be justified by something like the Principle of the Common Interest, but it seems to me difficult to maintain (without considerable stretching of the notion of a basic liberty) that in these cases basic liberties are being restricted only for the sake of the same or other basic liberties. It seems to me much more plausible and straightforward to say that in order to arrive at a policy in these cases we must balance the value of certain modes of exercise of a basic liberty not only against the exercise of other basic liberties but also against the enjoyment of other goods (uninterrupted sleep, undefaced public buildings, etc.). Something like this is surely true in the case of the restriction of expression by laws against defamation: different standards of defamation for, on the one hand, private, artistic or cultural expression and, on the other, political debate, seem to me obviously appropriate, and I take this to be the reflection of the differing values we place on the unfettered exercise of these forms of expression relative to, among other things, the value placed on safeguarding the primary good of self-respect.

One could of course maintain that what is balanced against liberty in these cases is not liberty itself but the *worth* of liberty. Since almost anything, including any significant increase or decrease in material well-being, can affect the worth of liberty, the general principle that basic liberties may be restricted only when this brings an increase (or is necessary to avoid a decrease) in the worth of the total system of basic liberties appears to be a weaker principle than Rawls wishes to defend. I suggest that Rawls' response here would be that while a great number of things can contribute to the worth of liberty, not every restriction of basic liberty that yields gains in other goods will yield sufficient gains to constitute a net increase in the worth of the total system of such liberties. This is what a restriction must do in order to be acceptable.

Conceivably, this principle can be made to fit the most obvious cases in which a restriction of basic liberty is justified. Given the rather diffuse character of the notion of "the worth of the total system of basic liberties," however, it is not a principle that is easy to apply. Under any account the decision as to when a restriction on basic liberty is justified will involve some difficult balancing, but I do not think that a clear guideline between acceptable and unacceptable balancing is obtained by describing everything in terms of "the worth of liberty." Such an approach might seem inviting if one thought that the notion of an increasing preference for basic liberties over other goods represented the most important theoretical element in the case for liberty. But I do not think that this is so. On the contrary it seems to me that the idea of an increasing preference for basic liberties leaves out or obscures the most important factors in the case for certain of the basic liberties, factors which Rawls' own discussion of these particular basic liberties brings out quite clearly.

2. Freedom of the Person

The argument from an increasing marginal preference for liberties over other primary goods is most appropriate as an account of the basis of freedom of the person. It is not completely clear from Rawls' discussion what this category of basic liberties is to encompass other than the protections against arrest and seizure embodied under "the rule of law," but I take it to include at least freedom of movement within the country and across its borders, freedom of choice in aspects of one's personal life, and perhaps also freedom from surveillance. The increasing preference for these liberties claimed by Rawls can be seen as deriving in part from the fact that they represent important conditions for the use and enjoyment of other goods. Beyond this, however, there is the fact that the interventions these liberties are intended to preclude constitute particularly deep intrusions into a person's life which anyone has strong reasons to want to avoid, both because of the real disruption they cause and because of their great symbolic impact.

We can of course imagine people who felt quite differently about these matters. To the extent that such differences are not merely the object of speculative imagination but the subject of real disagreement and controversy, the force of Rawls' argument for the priority of freedom of the person will be seriously weakened. But in such an event it seems clear that the case for these liberties will be genuinely in doubt. Rawls' analysis of the case for the freedoms of the person as a matter of relative preference thus seems quite appro-

priate; there is no obvious theoretical element in the case for these liberties that his analysis leaves out.

3. Liberties of Expression, Thought and Conscience

Freedom of speech and assembly, liberty of conscience and freedom of thought present a slightly different case. The argument for the priority of these liberties rests upon the recognition by the parties in the Original Position that as material conditions improve there will be a "growing insistence upon the right to pursue our spiritual and cultural interests."[42] As Rawls says in arguing for freedom of conscience, the parties "must assume that they may have moral, religious, or philosophical interests which they cannot put in jeopardy unless there is no alternative."[43]

Now this argument contains two distinguishable elements. The first is the recognition by the parties in the Original Position that, for the reasons discussed in connection with the argument against perfectionism, they cannot concede to the government any authority in matters of religious, moral or philosophic doctrine. As Rawls says,

> The government has no authority to render religious associations either legitimate or illegitimate any more than it has this authority in regard to art and science. These matters are simply not within its competence as defined by a just constitution. Rather, given the principles of justice, the state must be understood as the association consisting of equal citizens. It does not concern itself with philosophical and religious doctrine but regulates individuals' pursuit of their moral and spiritual interests in accordance with principles to which they themselves would agree in an initial situation of equality.[44]

The second element is the recognition by the parties that they will come to set a particularly high value on the pursuit of their "spiritual and cultural interests."

These two elements are clearly independent. To take the case of religion, the value that a group of people place on keeping their religious commitments will be reflected in such things as the amount of economic loss and disruption of the pattern of life they are willing to undergo to allow everyone to observe the holidays of his religion, attend services, etc. and in the lengths to which they are prepared

---

[42] *Id.* 543.
[43] *Id.* 206.
[44] *Id.* 212.

to go to recognize and respect the religious scruples of individual members against taking part in certain necessary tasks and activities. It is certainly possible that the cost a society is willing to bear in order to allow full freedom of religious observance might vary widely while the principle of the lack of governmental authority to decide between particular religious doctrines remained quite fixed.

This kind of variation in the value attached to religious observance, while possible, may in fact be unlikely if, as Rawls says, "[a]n individual recognizing religious and moral obligations regards them as binding absolutely in the sense that he cannot qualify his fulfillment of them for the sake of greater means for promoting his other interests."[45] This extraordinary importance attached to religious matters tends to overshadow the distinction I have tried to draw and makes it inviting to rest the case for toleration entirely on the claim that the parties in the Original Position can foresee that they will come to set an incomparably higher value on religious liberty (*i.e.* on the freedom to meet their religious commitments) than on other primary social goods. But this approach becomes less attractive if we think not only of religious liberty but of freedom of thought and expression more broadly construed. A society is apt to set rather different values on the fulfillment of religious commitments, the pursuit of scientific knowledge and the pursuit and enjoyment of excellence in the arts, and these differences will be reflected in the price the society is willing to bear in order to allow these activities to go forward. But in a society which recognizes freedom of thought and expression the regulation of these pursuits will be guided by a common principle that governments lack the authority to decide matters of moral, religious or philosophic doctrine (or of scientific truth) and hence also lack the authority to restrict certain activities on the grounds that they promulgate false or corrupting doctrines. Let me call this principle, which I have formulated only very crudely, the Principle of Limited Authority.

Taken alone such a principle does not constitute a complete doctrine of liberty or even of freedom of thought and expression. But it seems to me that this principle is the most important element in such a doctrine that can be established from the point of view of the Original Position. It is not possible to determine from that standpoint exactly what relative values are to be assigned to these pursuits and to other interests with which they may conflict. Nor is it possible to foresee from that standpoint what will be the best way of regulating

[45] *Id.* 207.

these pursuits so that they do not conflict. These are problems that can be dealt with only at a later stage when the full facts about a society and the preferences of its members are known. (I suspect that this process of balancing and coordination is what Rawls has in mind when he speaks of restricting particular basic liberties in order to strengthen the total system of basic liberties.) While it may be possible for the parties in the Original Position to foresee that in general they will attach a high value to their spiritual and cultural interests, such a general preference, or a resultant general principle that in the balancing process these liberties are to take precedence over other goods, seems to me to be less useful as the basis for a doctrine of freedom of thought and expression than the idea that the process of balancing must take place within the constraints imposed by something like the Principle of Limited Authority.

A doctrine of freedom of expression founded on this idea is suggested by Rawls on a number of occasions, in particular in his principle of the common interest, with its emphasis on the distinction between what might be called "neutral" and "non-neutral" grounds for restricting liberty. I think that some account of freedom of expression of this general type must be correct, although there are a number of difficulties in formulating such a view.[46] While I have some misgivings about Rawls' particular formulation (misgivings, *e.g.*, as to whether too much may be conceded to the doctrine of clear and present danger by his blanket allowance that liberty of conscience may be limited "when there is a reasonable expectation that not doing so will damage the public order which the government should maintain"[47]), it seems to me one of the strong points of Rawls' theory (as described in the first part of this section) that it provides a philosophical basis for an account of liberty of this type. It therefore seems to me important to ask whether this strength is adequately represented in his doctrine of the priority of liberty.

While it is not explicitly stated in the Priority Rule, the Principle of Limited Authority will be implied by clause (b) of that rule if (as seems plausible on the basis of the argument of the preceding section) we take governmental authority over matters of religion, etc. to represent an unequal liberty which would not be acceptable to those whose liberty is restricted. It is unclear, however, how this principle is related to the argument for the priority rule based on the increasing

---

[46] I have myself put forward a view of this kind in Scanlon, *A Theory of Freedom of Expression*, 1 PHIL. & PUB. AFFAIRS 204 (1972).

[47] RAWLS 213.

marginal value of liberty. There seem to be two possible interpretations of this argument.

While the parties in the Original Position might readily agree that there are conditions under which the pursuit of spiritual and cultural interests may be severely curtailed for the sake of other more pressing needs, it may seem unlikely, given the close relation between the Principle of Limited Authority and the conception of individual autonomy underlying the argument against perfectionism, that the parties would ever concede to a government the right to decide matters of moral, religious or philosophic doctrine. This suggests an interpretation of Rawls' argument according to which the Principle of Limited Authority applies under the General Conception of Justice as Fairness as well as under the Special Conception. What distinguishes the Special Conception, on this view, is just the increased importance that is attached to spiritual and cultural interests as the opportunity to pursue these interests presents itself and the demands of mere survival become less pressing. This interpretation is faithful to Rawls' description of the transition from the General Conception to the Special Conception as consisting of a shift in the ordering of primary social goods. But the Principle of Limited Authority is not a factor in this shift; it stands instead as a constant element of the theory. Given the importance of this principle from Rawls' point of view, it seems somewhat surprising on this interpretation that nothing resembling this principle is either stated or implied in Rawls' account of the General Conception.

An alternative, somewhat more extreme interpretation, and one which seems to me more likely to represent Rawls' view, would identify the Principle of Limited Authority as one of the distinguishing elements of the Special Conception. This means that there must be circumstances to which the General Conception of Justice as Fairness applies but in which the parties in the Original Position would not only allow the severe curtailment of expression on the grounds allowed under the Principle of the Common Interest but would also suspend the Principle of Limited Authority itself. I am not quite certain what such situations would be like. Presumably they would be situations in which cooperation on certain common tasks is not merely mutually advantageous but essential for survival or for the amelioration of intolerable conditions. If deep disagreements were to exist which made the basis of this cooperation fragile, and if close and uninterrupted cooperation were required to avoid consequences that would be disastrous for all, then perhaps it would be rational not only to accept rigid regulation on the time, place and manner of ex-

pression to prevent interference with essential work, but also to grant to the government the power to ban the expression of views likely to give rise to dangerous controversy or to dissention and doubt.

It seems to me most accurate to describe such situations as ones in which the circumstances of justice would be present only to a limited degree. Cooperation in certain tasks may be feasible and profitable and in these areas of common purpose considerations of justice may apply, dictating, *e.g.*, that the benefits and burdens of this cooperation (including liberties and constraints) should be shared in accordance with Rawls' Second Principle of Justice. But if the basis of this cooperation is quite shaky, and if the ends at which it aims are truly vital, then it might be rational for the parties involved to regard each other primarily as means to these ends. This attitude would be reflected, for example, in the parties' placing the smooth functioning of their institutions ahead of the right of individual members to raise and discuss with each other questions about the wisdom, viability or propriety of these institutions. I have some inclination to say that such a case would not represent cooperation on a footing of justice at all; collective actions to quell controversy in such circumstances are best seen not as the exercise of the distinctive authority of a just government in the sense defined by the Original Position, but rather as acts which must be justified on a case-by-case basis by appeal to the residual rights of the individuals involved to undertake those measures necessary to their self-defense and survival.[48]

However this may be, it is at least clear that the justification I have offered for limited tolerance in what might be called situations of partial justice depends upon the presence of conditions under which anything which undermines effective cooperation represents an immediate threat to all. When these conditions are lacking, such justification is also lacking, and, in addition, it becomes rational for people to seek to establish cooperation on a footing that gives full recognition to the status of the participants as autonomous equals, *i.e.* to something like Rawls' Special Conception.

One thing making this transition rational is the fact that under improving conditions individuals will develop religious, moral and philosophical interests and will want their institutions to safeguard their pursuit of these interests. But on the interpretation I have been discussing the Special Conception of Justice as Fairness can no longer be seen simply as what emerges under the General Conception once these interests begin to develop. For the transition to the Special

[48] Such a view is suggested in Scanlon, *supra* note 46, at 224-26.

Conception involves a fundamental change in the basis of cooperation, namely a move to what I called in the first part of this section cooperation on a footing of justice. Cooperation on this basis would be less apt to be rational for people if they did not place a high value on certain kinds of opportunity, but the defining elements of this form of cooperation go beyond this configuration of preferences, just as the defining elements of just cooperation in the economic sphere go beyond the structure of needs and interests that make such cooperation inviting.

III. CONTRA UTILITARIANISM

A dominant place in twentieth-century Anglo-American moral philosophy and, even more, in the normative thinking of American legal and social theorists, has been occupied by the family of theories called utilitarianism. Like the theories which, following Rawls, I have called perfectionist, these theories have a teleological structure. That is, they set out to define notions of *right* (the moral permissibility and impermissibility of actions, the justice or moral acceptability of social institutions) solely in terms of tendencies to promote certain specified ends (understood independently of considerations of right). In the case of classical utilitarianism as espoused by Bentham[49] this means that an action is held to be right (*i.e.* permissible) if and only if there is no alternative action available to the agent at the time which would yield greater net balance of pleasure over pain (or, alternatively, greater total happiness). Social institutions and pieces of legislation are to be judged similarly in terms of their tendency to promote this end. In determining the moral status of a course of action or an institution we are to take into account the happiness of every person affected, giving each equal weight.

It is useful to divide teleological theories into two groups on the basis of the end which they call upon us to promote. What I will call *hard* teleological theories are those which define notions of right in terms of the tendency to bring about certain states of affairs which are not states of consciousness in individuals and which are held to be valuable quite independently of their tendency to bring enjoyment, pleasure or satisfaction to individuals. *Soft* teleological theories, on the other hand, are those which take as the end to be promoted some state or states of consciousness of individuals, *e.g.* pleasurable sensations of some kind.

---

[49] *See* J. BENTHAM, INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION ch. 1 (1907).

Hard teleological theories (some or all of which are what Rawls calls perfectionist theories—I am not certain how broadly he intends this term to apply) leave open the possibility that an overall net sacrifice in the quality of the lives of all persons now and in the future could be justified if this were required for promotion of the specified goal. Thus, soft teleological theories, of which utilitarianism is the principal example, put themselves forward as a humanistic alternative. If the states of consciousness on which a utilitarian theory is based are ones which all human beings are capable (in roughly equal degrees) of experiencing, then that theory also has a democratic cast by comparison with perfectionist theories, in which all weight is given to the promotion of accomplishments of which only a few persons may be capable. Under utilitarianism everyone counts and is counted equally. This does not at once mean that everyone is to be treated equally, however, for distributive considerations per se are irrelevant in the utilitarian scheme, and it is quite conceivable that two quite different distributions of material goods might yield the same total of satisfactions and thus have equal claim to be promoted on utilitarian grounds. Natural empirical assumptions about the way in which satisfaction is in fact produced suggest, however, that equality of material circumstances has a special claim to the utilitarian's attention. The principal empirical assumption here is what is generally called the principle of diminishing marginal utility: the principle that each successive increment of a good produces a smaller increase in satisfaction than the preceding (equal) increment. If this is true for all of us and for all goods (hence generally for increments of real income), and if different persons' utility functions (*i.e.* the curve relating amounts of goods to satisfaction produced) are roughly similar, then every step away from equality of real income will be disadvantageous from a utilitarian point of view: the loss in satisfaction of the persons who receive a less than equal share will be greater than the gain in satisfaction for those whose real income is increased.

A parallel empirical argument provides a prima facie case for a kind of libertarianism on utilitarian grounds. If each person is the best judge of his own satisfactions, and if each person is naturally motivated to choose those alternatives which promise him greater satisfaction, then there is reason to think it good utilitarian policy to leave every person alone in those actions that concern primarily himself and affect others scarcely at all. It is worth pointing out that the motivational assumption made here must be plausible if the utilitarian theory itself is to be plausible. The main intuitive argument for the theory rests on the claim that what I have been calling satisfaction

represents for each person what is desirable in any event or course of action; the moral point of view, it is then claimed, just requires of us that we care equally about everyone else's fate in just the way we care about our own, *i.e.* that we seek to maximize the total of their (and our) satisfactions, counting each for one and none for more than one.

To a much greater degree than the other alternative views so far considered, utilitarianism can be formulated and given a plausible defense within Rawls' Original Position construction. Although I have so far been considering the classical version of utilitarianism in which it is the total of satisfactions that is to be maximized, it will be advantageous to focus on the arguments which can be offered in the Original Position for and against the principle of average utility, the principle which holds those social institutions to be morally preferable under which the average level of satisfactions is maximized.

What the proponent of the principle of average utility claims, according to Rawls,[50] is that each party in the Original Position should reason as follows:

> The worth to me of a given allotment of material goods and social opportunities (*i.e.* the satisfaction that allotment will bring me) will depend upon my tastes, desires and abilities. All I can say without knowledge of these is that I want to have that combination of, on the one hand, social goods and, on the other hand, tastes and abilities which will yield me the greatest level of satisfaction. In the absence of any information about my position in society or my tastes and abilities it is rational for me to assume that if I am a member of a particular society, the probability of my being a person in a certain social position is represented by the proportion of the total population who are in that position. Similarly, knowing nothing of my particular interests and talents, it is rational for me to suppose that the probability that I, if I am a person in a certain social position in a given society, will have or develop certain interests and talents is represented by the proportion of the people in that position in the society who have those interests and talents. It follows that the expected value for me of being some person (I know not which one) in a given society is represented by the average of the levels of satisfaction enjoyed by people in that society. Consequently, it is rational for me to choose as the fundamental principle by which the basic structure of my society is to be judged the principle that material goods and opportunities are to be distributed in such a way that the average level of satisfaction is maximized.

Rawls offers a number of arguments in support of his claim that

[50] RAWLS 163-66.

his own Two Principles (with the Second Principle interpreted by what he calls the Difference Principle) would be chosen over the principle of average utility by the parties in the Original Position. The argument which has received widest attention is based upon an appeal to what is known as the maximin criterion for rational choice under uncertainty. A "choice under uncertainty" in the technical sense in which this phrase is used in decision theory is a choice between alternatives some of which may produce a number of different outcomes depending upon circumstances which are beyond the chooser's control and to which he has no positive grounds for assigning probabilities. The maximin criterion specifies that in such a situation one should "maximize the minimum," *i.e.* seek to insure oneself the highest obtainable minimum by choosing that alternative whose worst outcome is at least as good as the worst outcome of any other alternative. Applied to the problem of the Original Position this means that the parties' main concern should be to maximize the expectations of the members of the worst-off class in their society. This leads them to choose as a principle of distributive justice what Rawls calls the Difference Principle.

The maximin criterion is an extremely conservative principle of choice. Faced with a choice between two alternatives, the first of which will yield us either one dollar or $100 depending on circumstances the relative likelihood of which we cannot determine, and the second of which will guarantee us two dollars in either case, the maximin criterion bids us choose the latter alternative. Rawls argues[51] that even though the maximin criterion may be too conservative to commend itself for general adoption, nonetheless the circumstances of the particular choice faced by the parties in the Original Position are such as to make a conservative principle of choice appropriate for anyone in that situation (and appropriate on objective grounds quite independent of subjective factors such as one's taste for gambling). These circumstances are, first, the fact that the principles chosen in the Original Position will determine the basic features of the society in which the parties live. What is at stake here is not one choice among many of approximately equal import but a choice of more fundamental importance than any other, one whose consequences are not likely to be erased or modified by subsequent choices or circumstances. In addition, given the assumption that prevailing conditions are ones which make social cooperation possible (we are not talking of two castaways fighting over one plank) it is presumably possible

[51] *Id.* 28.

in the society to which the parties belong to assure everyone of tolerable conditions of life. If this is so, then the parties to the Original Position, while they naturally have an interest in improving their condition above whatever level can be guaranteed for all, should rationally be more concerned to insure that they do not have to accept conditions below this level, which could be entirely unbearable.

Like the argument for the principle of average utility, this argument addresses itself to the interest which the parties to the Original Position have in doing as well for themselves in society as they can. Both arguments are thus examples of the reductive strategy which Rawls mentions as one of the primary strategies of the Original Position construction: to reduce the problem of determining principles of justice to a problem of individual prudential choice under uncertainty, a type of problem which has been extensively studied and one about which we may hope to have a clearer view. Rawls advances other arguments for the choice of his Two Principles, including arguments which appeal to other aims of the parties, *e.g.* their aims to select principles which it will be possible for them to uphold in the conditions of their actual society, to select principles which will support a stable social order, and to select principles which will express their nature as rational beings and provide a secure foundation for their self-respect. I will discuss some of these arguments in the following section. In the remainder of this section I want to consider in more detail the contrast between Rawls' principles and various forms of utilitarianism.

It is absolutely crucial to the argument presented above for the principle of average utility that a subjective quantity such as satisfaction rather than, say, primary social goods or some other objective index of real income be taken as the measure of relative well-being. As an alternative to be considered in the Original Position, a principle requiring the maximization of the average share of primary social goods would totally lack plausibility; for, since there will be wildly different distributions of such goods which yield the same *average*, the adoption of such a principle would represent an extreme form of gambling. Against the choice of this principle Rawls' argument for the rationality of conservatism in the choice of basic institutions seems exactly right. But the effect of the standard empirical assumptions (*i.e.* diminishing marginal utility) about how satisfaction is produced by material goods and opportunities is greatly to narrow the range of possible variation among those distributions of primary social goods which yield the same average level of *satisfaction*. Since these assumptions have quite a bit of intuitive plausibility for most people (perhaps

as much as the maximin criterion itself) this considerably undermines the force of Rawls' argument. This reliance on the special properties of satisfaction, however, exposes the utilitarian to another line of attack, one directed against the notion of satisfaction itself.

In accordance with the reductive strategy I have mentioned, what is appealed to in the argument for the principle of average utility is not what might be called the ethical role of the notion of satisfaction, but rather its psychological role in the motivation and explanation of the rational choices of a single self-interested individual. Here satisfaction (or in other formulations, happiness or pleasure) is supposed to represent that thing for the sake of which all other things are desired, and, since other things are desirable in proportion to the satisfaction etc. they yield, satisfaction serves as a common denominator which makes possible rational choice between apparently disparate alternatives (a day at the races, a night at the opera, an afternoon in bed). Rawls argues at some length that neither satisfaction nor happiness nor pleasure nor any other psychological quantity is a plausible candidate for the role of "dominant end," *i.e.* that thing for the sake of which all other things are to be desired. We do often speak of some things bringing us more satisfaction than others (or more pleasure or happiness) and of some courses of action turning out not to bring us satisfaction, but, Rawls argues, this should not be taken as a reference to some psychological quantity which underlies all our choices; rather, it is just another way of saying that certain things are to be preferred or that others do not, after all, seem to be preferable and that we are inclined to look for some alternative. Rawls also argues that there is no necessity to posit a dominant end in order to give an account of rational choice between disparate alternatives. It is simply a fact that we can make such choices without the aid of some psychological common standard. Rawls gives his own account of how some choices of this kind are possible, an account which I have briefly sketched in my discussion of the good for a person.[52] On Rawls' analysis, one method we may employ in choosing between disparate alternatives is to ask ourselves whether they do in fact answer to some of the same ends and, to the extent that this is so, which is preferable on these grounds. But this is not the whole process of rational choice; there are many other methods.

Although I find Rawls' arguments on these points entirely compelling and will suppose them to be correct in the discussion which follows, I cannot rehearse them in detail here. It should be noted,

[52] *See* Section II.A *supra.*

however, that Rawls' attack on the idea of a dominant end (satisfaction, pleasure or happiness) which is to be maximized in all rational self-interested choices is not an attack on the notion of utility as it is used in modern decision theory. In this theory (as to a large extent in Rawls' own account of rational choice) it is the notion of preference which is taken as primary. Decision theory uses the notion of a utility function to describe and systematize a person's preferences, but makes no claim to having explained those preferences by isolating that thing for the sake of which the preferred alternatives are to be preferred.

The idea that satisfaction represents that thing for the sake of which all other things are to be chosen is invoked in a particularly strong form in the argument for the principle of average utility. In this argument the particular tastes, desires, talents and abilities which are developed and exercised in a life, the choices which determine that life and the circumstances in which it is lived are all treated as quite incidental and secondary to the fundamental question of how much satisfaction the life promises to the person who lives it. The utilitarian chooser in the Original Position takes no facts about himself into account in making his choice of principles; he appraises alternatives not from the point of view of a person who will have some (at present unknown) system of ends, but from the point of view of a subject who will enjoy one quantity or another of satisfaction.

Now it may be said here that a person in the Original Position is prevented by the veil of ignorance from taking any facts about himself into account in making his choice of principles and therefore that the oddity I am objecting to is due to the idea of choice behind a veil of ignorance and not to utilitarianism. But there is more than one way of dealing with the lack of knowledge which those in the Original Position suffer. One possibility, the one just described, is to take the view that even if one does not know what one's ends and abilities are, the one thing one does know is that greater satisfaction is to be preferred to less, and this provides a way of ranking alternative lives which is neutral with respect to differences in tastes and abilities. Another possibility, however, is to consider those material goods and opportunities (the primary social goods) which, whether because they are necessary as means or for other reasons, are to be desired no matter what one's ends are. The latter alternative gives a priority to the choosing self which is lacking under the former. It represents the attitude: "Although I don't know what ends I, as a member of society, have, I know that my fundamental overall aim is to work out and pursue these ends. My purpose in choosing a conception of justice

should be to secure the best conditions for me to pursue this aim." The fundamental aim referred to here, while it might be said to represent the overall objective of a human life, is not a dominant end in the sense in which satisfaction claims to be; it is not that thing for which all other things—candy, *foie gras* and record albums—are chosen.

What I have here called the priority of the choosing self is of course the same notion on which the argument against perfectionism was found to hinge. This reflects the fact that the distinction between hard and soft teleological views is less deep than it at first appears. The aim of the move from hard to soft teleology was just to give a more fundamental place in ethical theory to each individual person. But once the notion of utility (satisfaction, pleasure, happiness, etc.) is identified with any particular complex of psychological states the activity of the rational self is again reduced to the status of a means to the production of these states. As Rawls says,

> The parties [to the Original Position] regard moral personality and not the capacity for pleasure and pain as the fundamental aspect of the self. They do not know what final aims persons have, and all dominant-end conceptions are rejected. Thus it would not occur to them to acknowledge the principle of utility in its hedonistic form. There is no more reason for the parties to agree to this criterion than to maximize any other particular objective.[53]

Just as the principle of utility as a criterion for first-person prudential choice leads a person to see his talents, ends and desires as means to the production of satisfaction, so, as an ethical principle, it requires a person to take the view that his life and activities have a claim on others for support and noninterference based solely on and strictly proportional to the amount of satisfaction they produce. Each person is to look at others (and himself) as so many lines along which resources may be allocated, the overall aim being to produce the greatest total (or average) satisfaction.

A utilitarian might respond that this outlook is not so implausible as I have tried to make it sound, and he might add that the rejection of this outlook is largely self-serving. "We all should admit," he might say, "that our more frivolous pursuits should not be provided for at the expense of enough food for others. Moreover, when we consider the natural explanation for this fact (based on the principle of diminishing marginal utility) we will see that it rests on an outlook strikingly similar to the one you have just caricatured."

---

[53] RAWLS 563.

But we must distinguish here between the intuitive judgment that more basic needs have a stronger claim to attention than less basic ones and the utilitarian explanation which finds the reason for this judgment in the fact that greater satisfaction is produced by fulfilling the former than by fulfilling the latter. Such an explanation is no more plausible here than in the case of prudential choice by a single individual: I believe that it is more important for me to have enough to eat than to have a new electric typewriter, but this fact is not to be explained by claiming that the former produces for me a greater quantity of that same thing (satisfaction) that makes the latter desirable. (In fact the claim that we choose between alernatives by comparing the amounts of satisfaction they offer seems to have its greatest plausibility when we are dealing with needs or wants of approximately equal urgency; it lacks appeal where the utilitarian needs it most: in the comparison of more basic needs with more frivolous ones.)

Now a utilitarian might respond at this point (if he has not done so earlier) that it is unfair to attack utilitarianism by attacking its hedonistic variant. Once we admit, as moral common ground, the idea that greater priority is to be attached to the fulfillment of some needs than to the fulfillment of others, how can we first assume that the utilitarian must go beyond this idea to an explanation in terms of a supposed psychological quantity, and then criticize him for doing this? Once a hierarchy of needs is admitted the utilitarian can base his theory directly on this hierarchy. Indeed, Rawls himself appeals to such a hierarchy in defending the use of the maximin rule in the Original Position when he points out that the parties should be more concerned with making sure that they do not fall below the level of well-being that can be guaranteed for all than with attempting to advance themselves above this level.

The modified utilitarian theory I have in mind here would be based on a hierarchy of levels of well-being measured in terms of some neutral standard such as Rawls' primary social goods. The goal of the theory (the standard by which acts and institutions are to be appraised) is to move as many people as high up in this ranking as possible, with the greatest importance to be set on increasing (and on not decreasing) the well-being of those who at a given time enjoy the lowest standard of well-being. The intuitive justification for this theory is just a depsychologized version of the intuitive argument for utilitarianism: each person attaches a positive value to increases in his own well-being (as measured by the standard in question), and a principle of decreasing marginal strength of preference applies here,

i.e. he attaches less importance to the difference between two positions "higher up" in the hierarchy of levels of well-being than to a quantitatively similar difference between positions "lower down" in the hierarchy. What the theory asks is that we take this same attitude towards the well-being of each person.

Now in order to apply this theory we must have some method for making decisions when faced with a choice between different complexes of gains to some people and losses to others. There is a problem here in specifying the terms in which the relevant comparisons are to be made without either introducing explicitly moral notions or else falling back on some psychological quantity, but the crucial point for the theory arises in determining how the choices are to be made on the basis of these comparisons, specifically how the bias in favor of helping (and not hurting) those who are less well off is to be expressed.

One alternative is to stick with the idea that while, in general, more basic needs are to take priority over less basic ones, nonetheless it is in principle possible that sufficiently large gains to enough people even at quite a high level of well-being could justify losses by a few people whose level of well-being is relatively low. On this alternative the basic teleological structure of utilitarianism is preserved—gains really are being balanced against losses—although it may be difficult or impossible to specify in substantive terms what it is that is being "maximized." A person who took this alternative might, in historic utilitarian fashion, maintain it as a rule of thumb (not a theoretical principle but a practical guide rarely to be broken) that the interests of those at lower levels of well-being should never be sacrificed for the benefit of those at higher levels.

A second alternative would be to make an initial theoretical decision based on the observation that more basic needs are not to be sacrificed to less basic ones and to maintain from the outset a rule against taking the gains of some as justification for the losses of others who are at a lower level of well-being. On this view it becomes unnecessary to compare in each case the gains of some with the losses of others, and we no longer have a theory which can be construed as calling upon us to maximize the sum (or average) over the population of any quantity at all. What we have instead is a theory directing that all decisions be made with the aim of increasing the well-being of those currently worst off. Thus, this apparently more dogmatic alternative brings us close to Rawls' Difference Principle. Moreover, if we take as a crucial step in Rawls' argument his appeal (mentioned above) to the asymmetry of the attitudes of the parties in the Original Position

to gains above the level of well-being that can be guaranteed to all and losses below this level, then the case he offers comes close to the form sketched at the beginning of this paragraph. Despite these similarities, however, there are significant differences between Rawls' position and the modified "utilitarian" position just described. I will discuss these differences more fully in the following section.

IV. Distributive Justice and the Difference Principle

Rawls is concerned with justice in only one of the many senses of the term. For him, questions of justice are questions of how the benefits and burdens of social cooperation are to be shared, and the principles of justice he develops are to apply in the first instance not to arbitrary distributions of goods but to the basic institutions of society which determine "the assignment of rights and duties and . . . regulate the distribution of social and economic advantages."[54] Rawls' principles apply to particular distributions only indirectly: a distribution may be called just if it is the result of just institutions working properly, but the principles provide no standard for appraising the justice of distributions independent of the institutions effecting them.[55] Conceived of in this way, principles of justice are analogous to a specification of what constitutes a fair gamble. If a gamble is fair then its outcome, whatever it may be, is fair and cannot be complained of. But the notion of a fair gamble provides no standard for judging particular distributions (Smith and Harris win five dollars, Jones loses ten dollars) as fair or unfair when these are considered in isolation from particular gambles which bring them about.

The principle which Rawls offers for appraising the distributive aspects of the basic structure of a society is his Second Principle of Justice which, considerations of the priority of liberty aside, is equivalent to what he calls the General Conception of Justice as Fairness. This principle is stated as follows: "Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity."[56]

According to clause (a) of this principle, which Rawls refers to as the Difference Principle, a system of social and economic inequalities is just only if there is no feasible alternative institution under which the expectations of the worst-off group would be greater. The phrase

---

[54] *Id.* 61.

[55] *Id.* 88.

[56] *Id.* 83. Rawls' final formulation of this principle, *id.* 312, incorporates considerations of justice between generations which the present discussion leaves aside.

"fair equality of opportunity" in clause (b) requires not only that no one be formally excluded from positions to which special benefits attach, but also that persons with similar talents and inclinations should have similar prospects of attaining these benefits "regardless of their initial place in the social system, that is, irrespective of the income class into which they are born."[57] The rationale behind this principle, particularly the motivation for clause (a), will be discussed at length below. First, however, I will consider briefly how the principle is to be applied.

A. *The Difference Principle and Its Application*

The most natural examples of inequalities to which Rawls' principle might be applied involve the creation of new jobs or offices to which special economic rewards are attached or an increase in the income associated with an existing job. But the intended application of the principle is much broader than this. It is to apply not only to inequalities in wealth and income but to all inequalities in primary social goods, *e.g.* to the creation of positions of special political authority. Further, its application is not limited to "jobs" or "offices" in the narrow sense but includes all the most general features of the basic structure of a society that give rise to unequal shares of primary social goods. In the case of economic goods these will include the system of money and credit, the laws of contract, the system of property rights and the laws governing the exchange and inheritance of property, the system of taxation, the institutions for the provision of public goods, etc.

It is fairly clear how Rawls' principle is to apply to the creation of one new office to which special rewards are attached (or to the assignment of new rewards to one existing position) in an otherwise egalitarian society: such an inequality is just only if those who do not directly benefit from this inequality by occupying the office benefit indirectly with the result that they too are better off than they were before (and than they would be if the benefits in question were distributed in any alternative way). It is less obvious how the principle is to apply in the more general case of complex institutions with many separable inequality-generating features. Rawls deals with this problem by specifying that institutions are to be appraised as a whole from the perspective of representative members of each relevant social position. The Difference Principle requires that the total system of inequalities be so arranged as to maximize the expectations of a representative member of the class which the system leaves worst off.

[57] *Id.* 73.

The notions of relevant social position and the expectations of a representative person in such a position require explanation. Relevant social positions in Rawls' sense are those places in the basic structure of society which correspond to the main divisions in the distribution of primary social goods. (He mentions the role of "unskilled worker" as constituting such a position.[58]) Rawls believes that the distribution of other primary social goods will be closely enough correlated with income and wealth that the latter can be taken as an index for identifying the least advantaged group. Accordingly, he suggests that the class of least advantaged persons may be taken to include everyone whose income is no greater than the average income of persons in the lowest relevant social position (or alternatively everyone with less than half the median income and wealth in the society[59]). To compute the expectations of a representative member of a given social position one takes the average of the shares of primary social goods enjoyed by persons in that position. Thus, while the parties in the Original Position do not estimate the value to them of becoming a member of a given society by taking the likelihood of their being a member of a particular social position to be represented by the proportion of the total population that is in that position, they do estimate the expected value (in primary social goods) of being a member of a particular social position by taking the likelihood that they will have any particular feature affecting the distribution of primary social goods within that position to be represented by the fraction of persons in the position who have that feature. Rawls does not explicitly discuss his reasons for allowing averaging within a social position when he has rejected it in the more general case. A more extreme position eschewing averaging would require maximizing the expectations of the worst-off individual in society. The Difference Principle occupies a position somewhere between this extreme and the principle of maximizing the average share of primary social goods across the society as a whole, its exact position within this range depending on how broadly or narrowly the relevant social positions are defined. The resort to averaging seems to some extent to be dictated by practical considerations: a coherent and manageable theory cannot take into account literally every position in a society.[60] In addition, the theoretical case against the use of averaging (as opposed to some more conservative method of choice) is weaker when we are concerned with differences in expectation within a single social position rather than differences between such positions. For here we

[58] *Id.* 98.
[59] *Id.*
[60] *Id.*

are not concerned with a single "gamble" with incomparably high stakes: intraposition differences are, by definition, limited, and each person's allotment is determined by a large number of independent factors, many of which are of approximately equal magnitude.[61]

There is a further problem about the notion of expectations which requires consideration. Rawls refers to the relevant social positions as "starting places," *i.e.* as the places in society people are born into.[62] Now the expectations of a person born into a family in a certain social position can be thought of as consisting of two components. First, there is the level of well-being he can expect to enjoy as a child. Presumably we may identify this with his parents' allotment of primary social goods. Second, there are his long term prospects as a member of society in his own right. If perfect fair equality of opportunity were attained then this latter component would not be substantially affected by the social and economic position of one's parents. As Rawls notes, however, such perfect equality of opportunity is unlikely, at least as long as the family is maintained,[63] so we may suppose that in general the second component will be heavily influenced by the first. One might conclude that the second component can be neglected entirely, reasoning that the distribution of social and economic advantages will influence the long term life prospects of a representative person born into the worst-off class mainly through its effect on the conditions in which such a person grows up. Taking this course would have the same consequences as deciding that what should be considered in applying the Difference Principle are not the expectations of a representative person born into the worst-off social position but the expectations of a representative person who winds up in that position after the social mechanism for assigning people to social roles has run its course.

But the principle which results from ignoring long term expectations seems to me unsatisfactory. Suppose we have a society in which there are 100 people in the lowest social position and twenty-five people in each of the two higher positions, and suppose it becomes known that the basic institutions of the society could be altered so that in later generations there would be fifty people in each of the three social positions, with the levels of wealth, income, authority, etc. associated with these positions remaining the same as they are now. Now it seems to me that a person in the lowest social position in this society is apt to be strongly in favor of this change. And such a person could plausibly support this preference by saying that the expectations of a repre-

[61] *Cf. id.* 169-71.
[62] *Id.* 96.
[63] *Id.* 74, 301.

sentative person born into his social position (in particular, the expectations of his children) would be better if this change were made than if it were not. This increase in expectations will not be captured by the interpretation of the Difference Principle just suggested or by any principle which focuses only on the levels of income, wealth, etc. associated with various positions in society while ignoring the way in which the population is distributed among these positions. Examples of this kind convince me that considerations of population distribution have to be incorporated in some way into Rawls' theory, and the most natural way to do this seems to me to be to bring them in through the notion of long term expectations.

But how is this to be done? The rule mentioned above that the expectations of a representative person in a given social position are to be determined by averaging the benefits enjoyed by persons in that position suggests that in a society with three relevant social positions whose average levels of income, wealth, authority, etc. can be indexed by p_1, p_2 and p_3, the long term prospects of a person born into the worst-off position should be represented by $a_1p_1 + a_2p_2 + a_3p_3$, where a_1, a_2 and a_3 are the fractions of people born into the worst-off position who wind up in each of the three places.

But the adoption of averaging as the method for computing long term expectations has unpleasant consequences for Rawls' theory. To the extent that the inequalities in childhood expectations resulting from the unequal economic and social positions of different families are eliminated (perhaps by eliminating the institution of the family itself), the first component in the expectations of a representative person will become the same for everyone regardless of the social position into which he is born, and Rawls' requirement that the expectations of a representative person in the lowest social position be maximized becomes the requirement that we maximize the second component of these expectations, *i.e.* the long term expectation $a_1p_1 + a_2p_2 + a_3p_3$. Moreover, to the extent that fair equality of opportunity is achieved (and barring the formation of a genetic elite) the coefficients a_1, a_2 and a_3 in this polynomial will become the same for every representative person regardless of social class, and the polynomial will thus come to express the average share of primary goods enjoyed by members of the society in question. It follows that on the interpretation just suggested Rawls' Difference Principle will be distinct from the principle requiring us to maximize the average share of primary social goods only so long as the inequalities resulting from the institution of the family persist or the fair equality of opportunity required by clause (b) of the Second Principle is otherwise not achieved. Even if

fair equality of opportunity is an unattainable ideal this conclusion seems to me unacceptable for Rawls' theory. As was pointed out above,[64] the principle of maximum average primary social goods is an extremely implausible one, much less plausible than the principle of maximum average utility. I see no reason to think that this principle would be acceptable even if perfect equality of opportunity were to obtain.

The problem here is how to give some weight to the way in which the population is distributed across social positions without introducing aggregative considerations in such a way that they take over the theory altogether (or would do so but for the "friction" introduced by imperfect equality of opportunity). One way of dealing with this problem which seems to me in the spirit of Rawls' theory would be to modify the Difference Principle to require the following:

> First maximize the income, wealth, etc. of the worst-off representative person, then seek to minimize the number of people in his position (by moving them upwards); then proceed to do the same for the next worst-off social position, then the next and so on, finally seeking to maximize the benefits of those in the best-off position (as long as this does not affect the others).[65]

This seems to me a natural elaboration of what Rawls calls the Lexical Difference Principle.[66] It also has the advantage of dealing with the problem of population distribution without introducing the summing or averaging of benefits across relevant social positions. There are obviously many variations on this theme as well as many altogether different approaches.[67]

B. *The Argument for the Difference Principle*

I return now to the central question of the rationale behind the Difference Principle. The intuitive idea here is that a system of in-

---

[64] *See* p. 1050 *supra.*

[65] This solution was suggested to me by Bruce Ackerman.

[66] RAWLS 83. *See* text accompanying note 68 *infra.*

[67] One would be to take the position a person is "born into" to be defined not only by the social and economic status of his family but also by his inborn talents and liabilities, *i.e.* those features which will enable him to prosper in the society or prevent him from doing so. Given this definition of the "starting places," one could employ averaging as a method for representing the long term expectations of a representative person born into the worst-off such place without fear that the theory would collapse into the doctrine of maximum average primary social goods if the institution of the family were eliminated. Modifying the Difference Principle in this way would bring Rawls closer (perhaps too close) to what he calls "the principle of redress," the principle that the distribution of social advantages must be arranged to compensate for undeserved inequalities such as the inequalities of birth and natural endowment. *See* RAWLS 100-02.

equalities is just only if we can say to each person in the society, "Eliminating the advantages of those who have more than you would not enable us to improve the lot of any or all of the people in your position (or beneath it). Thus it is unavoidable that a certain number of people will have expectations no greater than yours, and no unfairness is involved in your being one of these people." The requirement that we be able to say this to *every* member of society, and not just to those in the worst-off group, corresponds to what Rawls calls the Lexical Difference Principle:

> [I]n a basic structure with n relevant representatives, first maximize the welfare of the worst-off representative man; second, for equal welfare of the worst-off representative, maximize the welfare of the second worst-off representative man, and so on until the last case which is, for equal welfare of all the preceding n-1 representatives, maximize the welfare of the best-off representative man.[68]

This form of the principle is called "lexical" since "lexical priority" is given to the expectations of the worse-off: the fate of the second worst-off group is considered only to decide between arrangements which do equally well for the worst-off, and so on for the higher groups, working always from the bottom up. This asymmetry of concern in favor of the worse-off is a central feature of the theory. Rawls remarks a number of times in contrasting his theory with utilitarianism that under the Difference Principle no one is "expected . . . to accept lower prospects of life for the sake of others."[69] But what this means, as Rawls himself notes,[70] is that no one is expected to take *less than others receive* in order that the others may have a greater share. It seems likely, however, that those who are endowed with talents which are much in demand will receive less in a society governed by Rawls' Difference Principle than they would if allowed to press for all they could get on a free market. Thus, in a Rawlsian society these people will be asked to accept less than they might otherwise have had, and there is a clear sense in which they will be asked to accept these

---

[68] *Id.* 83. I will regard this as the canonical formulation of Rawls' principle. When this version of the principle is fulfilled there is a clear sense in which prevailing inequalities are "to everyone's advantage" since there is no one who would benefit from their removal. Fulfillment of the simple Difference Principle (that inequalities must benefit the worst-off) insures fulfillment of the lexical principle only if expectations of members of the society are " close knit"—it is impossible to alter the expectations of one representative person without affecting the expectations of every other representative person—and "chain connected"—if an inequality favoring group *A* raises the expectations of the worst-off representative person *B* then it also raises the expectations of every representative person between *B* and *A*. *Id.* 80-82.

[69] *Id.* 178, 180.

[70] *Id.* 103.

smaller shares "for the sake of others." What, then, can be said to these people?

Rawls' stated answer to this question consists in pointing out that the well-being of the better endowed, no less than that of the other members of society, depends on the existence of social cooperation, and that they can "ask for the willing cooperation of everyone only if the terms of the scheme are reasonable."[71] The Difference Principle, Rawls holds, represents the most favorable basis of cooperation the well-endowed could expect others to accept. Taken by itself this does not seem an adequate response to the complaint of the better endowed, for the question at issue is just what terms of cooperation are "reasonable."

The particular notion of "reasonable terms" that Rawls is appealing to here is one that is founded in the conception of social cooperation which he is propounding. The basis of this conception lies not in a particular bias in favor of the less advantaged but in the idea that economic institutions are reciprocal arrangements for mutual advantage in which the parties cooperate on a footing of equality. Their cooperative enterprise may be more or less efficient depending on the talents of the members and how fully these are developed, but since the value of these talents is something that is realized only in cooperation the benefits derived from these talents are seen as a common product on which all have an equal claim. Thus Rawls says of his Two Principles that they "are equivalent . . . to an undertaking to regard the distribution of natural abilities as a collective asset so that the more fortunate are to benefit only in ways that help those who have lost out."[72]

This same notion of the equality of the parties in a cooperative scheme is invoked in the following intuitive argument for the Difference Principle.

> Now looking at the situation from the standpoint of one person selected arbitrarily, there is no way for him to win special advantages for himself. Nor, on the other hand, are there grounds for his acquiescing in special disadvantages. Since it is not reasonable for him to expect more than an equal share in the division of social goods, and since it is not rational for him to agree to less, the sensible thing for him to do is to acknowledge as the first principle of justice one requiring an equal distribution. Indeed, this principle is so obvious that we would expect it to occur to anyone immediately.

---

[71] *Id.*

[72] *Id.* 179.

> Thus, the parties start with a principle establishing equal liberty for all, including equality of opportunity, as well as an equal distribution of income and wealth. But there is no reason why this acknowledgment should be final. If there are inequalities in the basic structure that work to make everyone better off in comparison with the benchmark of initial equality, why not permit them?[73]

If one accepts equality as the natural first solution to the problem of justice then this argument strongly supports the conclusion that the Difference Principle marks the limit of acceptable inequality. More surprisingly, it also appears to show (whether or not one accepts equality as a first solution) that the Difference Principle is the most egalitarian principle it would be rational to adopt. It is of course a difficult empirical question how much inequality in income and wealth the Difference Principle will in fact allow, *i.e.* how many economic inequalities will be efficient enough to "pay their own way" as the principle requires. The only theoretical limitation on such inequalities provided by Rawls' theory appears to be the possibility that glaring inequalities in material circumstances may give rise to (justified) feelings of loss of self-respect[74] on the part of those less advantaged, offsetting the material gains these inequalities bring them. One can thus make the Difference Principle more (or less) egalitarian by introducing a psychological premise positing greater (or lesser) sensitivity to perceived inequality. But as far as I am able to determine there is no plausible principle which is distinct from the Difference Principle and intermediate between it and strict equality. Since the inequalities allowed by the Difference Principle, while not great, may nonetheless be significant, this strikes me as a suprising fact. What it shows, perhaps, is that if one wishes to defend a position more egalitarian than Rawls' then one must abandon distributive justice as the cardinal virtue of social institutions, *i.e.* one must abandon the perspective which takes as the dominant moral problem of social cooperation that of justifying distributive institutions to mutually disinterested persons

---

[73] *Id.* 150-51.

[74] Inequalities give rise to loss of self-respect in Rawls' sense to the extent that they give a person reason for lack of confidence in his own worth and in his abilities to carry out his life plans. *Id.* 535. Whether given inequalities have this effect will depend not only on their magnitude but also on the public reasons offered to justify them. Rawls believes that effects of this kind will not be a factor in a society governed by the Difference Principle since the inequalities in wealth and income in such a society will not be extreme and will "probably [be] less than those that have often prevailed." *Id.* 536. In addition, the justification offered for those inequalities that do prevail will be one which supports the self-esteem of the less advantaged since this justification must appeal to the tendency of these inequalities to advance their good.

each of whom has a fundamental interest in receiving the greatest possible share of the distributed goods.[75]

The ideal of social cooperation which Rawls presents is naturally contrasted with two alternative conceptions of justice. The first of these is what Rawls calls the system of natural liberty.[76] This conception presupposes background institutions which guarantee equal liberties of citizenship in the sense of the First Principle and preserve formal equality of opportunity, *i.e.* "that all have at least the same legal rights of access to all advantaged social positions."[77] But no effort is made to compensate for the advantages of birth, *i.e.* of inherited wealth. Against the background provided by these institutions individuals compete in a free market and are free to press upon one another whatever competitive advantages derive from their different abilities and circumstances.

The second alternative is that of utilitarianism, understood broadly to include the two modified views presented at the end of the last section. The last of those views differed from the versions of utilitarianism criticized by Rawls in that it incorporated Rawls' principle that no one may be asked to accept a less than equal share in order that some others may enjoy correspondingly greater benefits. But even though it is not simply a maximizing conception, this view is like other forms of utilitarianism in holding it to be the duty of each person to make the greatest possible contribution to the welfare of mankind. Any asset one may have control over, whether a personal talent or a transferable good, one is bound to disburse in such a way as to make the greatest contribution to human well-being.[78] Utilitarianism is in this sense an asocial view; the relation taken as fundamental by the theory is that which holds between any two people when one has the capacity to aid the other. Relations between persons deriving from their position in common institutions, *e.g.* institutions of production and exchange, are in themselves irrelevant. It would be possible to maintain a view of this kind which focused only on the well-being of

---

[75] A position of this kind was put forward, for example, by Kropotkin. *See* P. KROPOTKIN, THE CONQUEST OF BREAD 62, ch. 13, *et passim* (Penguin ed. 1972). Kropotkin holds that if one accepts, as Rawls appears to, the view that the productive capacities of a society must be seen as the common property of its members, then one must reject the idea of wages (or any other way of tying distribution to social roles). Rather, the social product is to be held in common and used to provide facilities which meet the basic needs of all.

[76] RAWLS 72.

[77] *Id.*

[78] This aspect of utilitarianism is most clearly emphasized by William Godwin. *See* 2 W. GODWIN, ENQUIRY CONCERNING POLITICAL JUSTICE bk. VIII (3d ed. 1797) (facsim. ed. F. Priestley 1946).

members of a particular society, but such a restriction would appear arbitrary. The natural tendency of utilitarian theories is to be global in their application.

Rawls' Difference Principle can be seen as occupying a position intermediate between these two extremes. Like the system of natural liberty and unlike utilitarianism, Rawls' conception of justice applies only to persons who are related to one another under common institutions. The problem of justice arises, according to Rawls, for people who are engaged in a cooperative enterprise for mutual benefit, and it is the problem of how *the benefits of their cooperation* are to be shared. What the parties in a cooperative scheme owe one another as a matter of justice is an equitable share of this social product, and neither the maximum attainable level of satisfaction nor the goods and services necessary, given their needs and disabilities, to bring them up to a certain level of well-being.

The qualification "as a matter of justice" is essential here since justice, central though it is, is not the only moral notion for Rawls, and other moral notions take account of need and satisfaction in a way that justice does not. Rawls speaks, for example, of the duty of mutual aid, "the duty of helping another when he is in need or jeopardy, provided that one can do so without excessive risk or loss to oneself."[79] Now it seems likely that those to whom we are bound by ties of justice will fare better at our hands (or at least have a stronger claim on us) than those to whom we owe only duties of mutual aid; for justice, which requires that our institutions be arranged so as to maximize the expectations of the worst-off group in our society, says nothing about others elsewhere with whom we stand in no institutional relation but who may be worse off than anyone in our society. If this is so, then it may make a great deal of difference on Rawls' theory where the boundary of society is drawn. Are our relations with the people of South Asia, for example (or the people in isolated rural areas of our own country), governed by considerations of justice or only by the duties which hold between any one human being and another? The only satisfactory solution to this problem seems to me to be to hold that considerations of justice apply at least wherever there is systematic economic interaction; for whenever there is regularized commerce there is an institution in Rawls' sense, *i.e.* a public system of rules defining rights and duties etc.[80] Thus the Difference Principle

[79] RAWLS 114.
[80] *Id.* 55.

would apply to the world economic system taken as a whole as well as to particular societies within it.

In distinguishing justice from altruism and benevolence and taking it to apply only to arrangements for reciprocal advantage Rawls' theory is like the system of natural liberty. But a proponent of natural liberty takes "arrangements for reciprocal advantage" in the relevant sense to be arrangements arising out of explicit agreements. Such arrangements are just if they were in fact freely agreed to by the parties involved, and the background institutions of the system of natural liberty are designed to ensure justice in this sense. Since Rawls' Difference Principle constrains people to cooperate on terms other than those they would arrive at through a process of free bargaining on the basis of their natural assets, it is to be rejected. As Rawls says, the terms of this principle are equivalent to an undertaking to regard natural abilities as a common asset, and a proponent of natural liberty would say, I believe, that the terms of the principle apply only where such an undertaking has in fact been made.

Rawls holds, on the other hand, that one is born into a set of institutions whose basic structure largely determines one's prospects and opportunities. Background institutions of the kind described in the system of natural liberty are one example of such institutions; the various institutions satisfying the Difference Principle are another. Within the framework of such institutions one may enter into specific contractual arrangements with others, but these institutions themselves are not established by explicit agreement; they are present from birth and their legitimacy must have some other foundation. The test of legitimacy which Rawls proposes is, of course, the idea of hypothetical contract, as it is embodied in his Original Position construction.

The argument sketched here is obviously parallel to a familiar controversy about the bases of political obligation. The doctrine of natural liberty corresponds to the doctrine which seeks to found all political ties on explicit consent, and seems to me to inherit many of the problems of that view. For Rawls, on the other hand, the legitimacy of both political and economic institutions is to be analyzed in terms of a merely hypothetical agreement. (Indeed, Rawls does not separate the two cases.) The parallel between the problems of political institutions and those of economic institutions is often obscured because the political problem is thought of in terms of *obligation* while economic justice is thought of in terms of *distribution*.[31] But economic institu-

[31] For a discussion of political obligation relevant to economic contribution as well, see M. WALZER, OBLIGATIONS (1970).

tions, no less than political ones, must be capable of generating obligations, *viz.* obligations to cooperate on the terms these institutions provide in order to produce the shares to which others are entitled.[82]

The idea of such economic obligations raises a number of interesting issues which I can only mention here. Such an obligation to contribute would be violated, *e.g.*, by a person who, while wishing to receive benefits derived from the participation of others in a scheme of cooperation satisfying the Difference Principle, refused to contribute his own skills on the same terms, holding out for a higher level of compensation than the scheme provided. Presumably obligations of this kind do not in general prevent a person from opting out of a scheme of economic cooperation, any more than political obligation constitutes a general bar to emigration; but this does not mean (in either case) that people are always free to simply pick up and go. Further, there obviously are limits to what a just scheme can demand of those born into it and limits to how far their freedom to choose among different forms of contribution can be restricted. It seems likely that these limits would be defended, on Rawls' view, by appeal to an increasing marginal preference for "economic liberty" relative to other goods.

As I have argued above, the central thesis underlying the Difference Principle is the idea that the basic institutions of society are a cooperative enterprise in which the citizens stand as equal partners. This notion of equality is reflected in Rawls' particular Original Position construction in the fact that the parties are prevented by the veil of ignorance and the requirement that the principles they choose be general (*i.e.* contain no proper names or token reflexives) from framing principles which ensure them special advantages.[83] But the fact that it would be chosen under these conditions is not a conclusive argument for the Difference Principle since a person who favored the system of natural liberty would undoubtedly reject the notion that principles of justice must be chosen under these particular constraints. The situation here is similar to that of the argument against perfectionism: Rawls' defense of the Difference Principle must proceed in the main by setting out the ideal of social cooperation of which this prin-

---

[82] *See* RAWLS 313. The contribution side of the problem of economic justice is forcefully emphasized in R. NOZICK, ANARCHY, STATE AND UTOPIA (1974) (forthcoming). Nozick criticizes Rawls from the perspective of a purely contractarian view much more sophisticated and subtle than the system of natural liberty I have crudely described here.

[83] These considerations alone, of course, do not ensure that the parties in the Original Position will arrive at a principle of equal distribution even as a first solution. Given that they have no way to ensure a larger share for themselves the question remains whether they should settle for the maximin solution represented by the Difference Principle or gamble on receiving a larger share under some other rule.

ciple is the natural expression. The advantages of this ideal—*e.g.* the fact that institutions founded on this ideal support the self-esteem of their members and provide a public expression of their respect for one another—can be set out, and its ability to account for our considered judgments of justice can be demonstrated, but in the end the adoption of an alternative view is not wholly precluded. A person who, finding that he has valuable talents, wishes to opt for the system of natural liberty is analogous to the person who, knowing his own conception of the good, prefers a perfectionist system organized around this conception to what I have called "cooperation on a footing of justice." In both cases one can offer reasons why cooperation with others on a basis all could agree to in a situation of initial equality is an important good, but one cannot expect to offer arguments which meet the objections of such a person and defeat them on their own grounds.

I do not regard this residual indeterminacy as a failing of Rawls' book or as a source for skepticism. The conception of justice which Rawls describes has an important place in our thought, and to have presented this conception as fully and displayed its deepest features as clearly as Rawls has done is a rare and valuable accomplishment. Almost no one will read the book without finding himself strongly drawn to Rawls' view at many points, and even those who do not share Rawls' conclusions will come to a deeper understanding of their own views as a result of his work.