KANTIAN CONSTRUCTIVISM IN MORAL THEORY *

RATIONAL AND FULL AUTONOMY

IN these lectures I examine the notion of a constructivist moral conception, or, more exactly, since there are different kinds of constructivism, one Kantian variant of this notion. The variant I discuss is that of justice as fairness, which is presented in my book *A Theory of Justice*.[1] I have two reasons for doing this: one is that it offers me the opportunity to consider certain aspects of the conception of justice as fairness which I have not previously emphasized and to set out more clearly the Kantian roots of that conception. The other reason is that the Kantian form of constructivism is much less well understood than other familiar traditional moral conceptions, such as utilitarianism, perfectionism, and intuitionism. I believe that this situation impedes the advance of moral theory. Therefore, it may prove useful simply to explain the distinctive features of Kantian constructivism, to say what it is, as illustrated by justice as fairness, without being concerned to defend it. To a degree that it is hard for me to estimate, my discussion assumes some

515

acquaintance with *A Theory of Justice*, but I hope that, for the most part, a bare familiarity with its main intuitive ideas will suffice; and what these are I note as we proceed.

I would like to think that John Dewey, in whose honor these lectures are given, would find their topic hospitable to his concerns. We tend to think of him as the founder of a characteristically American and instrumental naturalism and, thus, to lose sight of the fact that Dewey started his philosophical life, as many did in the late nineteenth century, greatly influenced by Hegel; and his genius was to adapt much that is valuable in Hegel's idealism to a form of naturalism congenial to our culture. It was one of Hegel's aims to overcome the many dualisms which he thought disfigured Kant's transcendental idealism, and Dewey shared this emphasis throughout his work, often stressing the continuity between things that Kant had sharply separated. This theme is present particularly in Dewey's early writings, where the historical origins of his thought are more in evidence.[2] In elaborating his moral theory along somewhat Hegelian lines, Dewey opposes Kant, sometimes quite explicitly, and often at the same places at which justice as fairness also departs from Kant. Thus there are a number of affinities between justice as fairness and Dewey's moral theory which are explained by the common aim of overcoming the dualisms in Kant's doctrine.

I

What distinguishes the Kantian form of constructivism is essentially this: it specifies a particular conception of the person as an element in a reasonable procedure of construction, the outcome of which determines the content of the first principles of justice. Expressed another way: this kind of view sets up a certain procedure of construction which answers to certain reasonable requirements, and within this procedure persons characterized as rational agents of construction specify, through their agreements, the first principles of justice. (I use 'reasonable' and 'rational' to express different notions throughout, notions which will be explained below, in section v, 528–530.) The leading idea is to establish a suitable connection between a particular conception of the person and first principles of justice, by means of a procedure of construction. In a Kantian view the conception of the person, the procedure, and the first prin-

---

[2] See, for example, Dewey's *Outlines of a Critical Theory of Ethics* (1891) and *The Study of Ethics: A Syllabus* (1894) reprinted in *John Dewey: The Early Works*, 1882–1898 (Carbondale: Southern Illinois University Press, 1971), in volumes 3 and 4, respectively. From Dewey's critique of Kant in *Outlines*, pp. 290–300, and his statement of his own form of the self-realization doctrine, pp. 300–327, Dewey's debt to idealism is plain enough.

ciples must be related in a certain manner—which, of course, admits of a number of variations. Justice as fairness is not, plainly, Kant's view, strictly speaking; it departs from his text at many points. But the adjective 'Kantian' expresses analogy and not identity; it means roughly that a doctrine sufficiently resembles Kant's in enough fundamental respects so that it is far closer to his view than to the other traditional moral conceptions that are appropriate for use as benchmarks of comparison.

On the Kantian view that I shall present, conditions for justifying a conception of justice hold only when a basis is established for political reasoning and understanding within a public culture. The social role of a conception of justice is to enable all members of society to make mutually acceptable to one another their shared institutions and basic arrangements, by citing what are publicly recognized as sufficient reasons, as identified by that conception. To succeed in doing this, a conception must specify admissible social institutions and their possible arrangements into one system, so that they can be justified to all citizens, whatever their social position or more particular interests. Thus, whenever a sufficient basis for agreement among citizens is not presently known, or recognized, the task of justifying a conception of justice becomes: how can people settle on a conception of justice, to serve this social role, that is (most) reasonable for them in virtue of how they conceive of their persons and construe the general features of social cooperation among persons so regarded?

Pursuing this idea of justification, we take our examination of the Kantian conception of justice as addressed to an impasse in our recent political history; the course of democratic thought over the past two centuries, say, shows that there is no agreement on the way basic social institutions should be arranged if they are to conform to the freedom and equality of citizens as moral persons. The requisite understanding of freedom and equality, which is implicit in the public culture of a democratic society, and the most suitable way to balance the claims of these notions, have not been expressed so as to meet general approval. Now a Kantian conception of justice tries to dispel the conflict between the different understandings of freedom and equality by asking: which traditionally recognized principles of freedom and equality, or which natural variations thereof, would free and equal moral persons themselves agree upon, if they were fairly represented solely as such persons and thought of themselves as citizens living a complete life in an on-going society? Their agreement, assuming an agreement would be reached, is

conjectured to single out the most appropriate principles of freedom and equality and, therefore, to specify the principles of justice.

An immediate consequence of taking our inquiry as focused on the apparent conflict between freedom and equality in a democratic society is that we are not trying to find a conception of justice suitable for all societies regardless of their particular social or historical circumstances. We want to settle a fundamental disagreement over the just form of basic institutions within a democratic society under modern conditions. We look to ourselves and to our future, and reflect upon our disputes since, let's say, the Declaration of Independence. How far the conclusions we reach are of interest in a wider context is a separate question.

Hence, we should like to achieve among ourselves a practicable and working understanding on first principles of justice. Our hope is that there is a common desire for agreement, as well as a sufficient sharing of certain underlying notions and implicitly held principles, so that the effort to reach an understanding has some foothold. The aim of political philosophy, when it presents itself in the public culture of a democratic society, is to articulate and to make explicit those shared notions and principles thought to be already latent in common sense; or, as is often the case, if common sense is hesitant and uncertain, and doesn't know what to think, to propose to it certain conceptions and principles congenial to its most essential convictions and historical traditions. To justify a Kantian conception within a democratic society is not merely to reason correctly from given premises, or even from publicly shared and mutually recognized premises. The real task is to discover and formulate the deeper bases of agreement which one hopes are embedded in common sense, or even to originate and fashion starting points for common understanding by expressing in a new form the convictions found in the historical tradition by connecting them with a wide range of people's considered convictions: those which stand up to critical reflection. Now, as I have said, a Kantian doctrine joins the content of justice with a certain conception of the person; and this conception regards persons as both free and equal, as capable of acting both reasonably and rationally, and therefore as capable of taking part in social cooperation among persons so conceived. In addressing the public culture of a democratic society, Kantian constructivism hopes to invoke a conception of the person implicitly affirmed in that culture, or else one that would prove acceptable to citizens once it was properly presented and explained.

I should emphasize that what I have called the "real task" of

justifying a conception of justice is not primarily an epistemological problem. The search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society replaces the search for moral truth interpreted as fixed by a prior and independent order of objects and relations, whether natural or divine, an order apart and distinct from how we conceive of ourselves. The task is to articulate a public conception of justice that all can live with who regard their person and their relation to society in a certain way. And though doing this may involve settling theoretical difficulties, the practical social task is primary. What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us. We can find no better basic charter for our social world. Kantian constructivism holds that moral objectivity is to be understood in terms of a suitably constructed social point of view that all can accept. Apart from the procedure of constructing the principles of justice, there are no moral facts. Whether certain facts are to be recognized as reasons of right and justice, or how much they are to count, can be ascertained only from within the constructive procedure, that is, from the undertakings of rational agents of construction when suitably represented as free and equal moral persons. (The points noted in this paragraph will be discussed in more detail in the third lecture.)

II

These first remarks were introductory and intended merely to suggest the themes of my discussion. To proceed, let's specify more exactly the above-mentioned impasse in our political culture as follows, namely, as a conflict between two traditions of democratic thought, one associated with Locke, the other with Rousseau. Using the distinction drawn by Benjamin Constant between the liberties of the moderns and the liberties of the ancients, the tradition derived from Locke gives pride of place to the former, that is, to the liberties of civic life, especially freedom of thought and conscience, certain basic rights of the person, and of property and association; while the tradition descending from Rousseau assigns priority to the equal political liberties and values of public life, and views the civic liberties as subordinate. Of course, this contrast is in many respects artificial and historically inaccurate; yet it serves to fix ideas and enables us to see that a mere splitting of the difference between these two traditions (even if we should agree on a favored interpre-

tation of each) would be unsatisfactory. Somehow we must find a suitable rendering of freedom and equality, and of their relative priority, rooted in the more fundamental notions of our political life and congenial to our conception of the person.

But how are we to achieve this? Justice as fairness tries to uncover the fundamental ideas (latent in common sense) of freedom and equality, of ideal social cooperation and of the person, by formulating what I shall call "model-conceptions." We then reason within the framework of these conceptions, which need be defined only sharply enough to yield an acceptable public understanding of freedom and equality. Whether the doctrine that eventually results fulfills its purpose is then decided by how it works out: once stated, it must articulate a suitable conception of ourselves and of our relation to society, and connect this conception with workable first principles of justice, so that, after due consideration, we can acknowledge the doctrine proposed.

Now the two basic model-conceptions of justice as fairness are those of a *well-ordered society* and of a *moral person*. Their general purpose is to single out the essential aspects of our conception of ourselves as moral persons and of our relation to society as free and equal citizens. They depict certain general features of what a society would look like if its members publicly viewed themselves and their social ties with one another in a certain way. The *original position* is a third and mediating model-conception: its role is to establish the connection between the model-conception of a moral person and the principles of justice that characterize the relations of citizens in the model-conception of a well-ordered society. It serves this role by modeling the way in which the citizens in a well-ordered society, viewed as moral persons, would ideally select first principles of justice for their society. The constraints imposed on the parties in the original position, and the manner in which the parties are described, are to represent the freedom and equality of moral persons as understood in such a society. If certain principles of justice would indeed be agreed to (or if they would belong to a certain restricted family of principles), then the aim of Kantian constructivism to connect definite principles with a particular conception of the person is achieved.

For the present, however, I am concerned with the parties in the original position only as rationally autonomous agents of construction who (as such agents) represent the aspect of rationality, which is part of the conception of a moral person affirmed by citizens in a well-ordered society. The rational autonomy of the parties in the

original position contrasts with the full autonomy of citizens in society. Thus *rational* autonomy is that of the parties as agents of construction: it is a relatively narrow notion, and roughly parallels Kant's notion of hypothetical imperatives (or the notion of rationality found in neo-classical economics); *full* autonomy is that of citizens in everyday life who think of themselves in a certain way and affirm and act from the first principles of justice that would be agreed to. In section v, I shall discuss the constraints imposed on the parties which enable the original position to represent the essential elements of full autonomy.

Let us briefly recall the features of a well-ordered society most relevant here.[3] First, such a society is effectively regulated by a public conception of justice; that is, it is a society in which every one accepts, and knows that others likewise accept, the same first principles of right and justice. It is also the case that the basic structure of society, the arrangement of its main institutions into one social scheme, actually satisfies, and is believed by all on good grounds to satisfy, these principles. Finally, the public principles of justice are themselves founded on reasonable beliefs as established by the society's generally accepted methods of inquiry; and the same is true of the application of these principles to judge social institutions.

Second, the members of a well-ordered society are, and view themselves and one another in their political and social relations (so far as these are relevant to questions of justice) as, free and equal moral persons. Here there are three distinct notions, specified independently: freedom, equality, and moral (as applied to) person. The members of a well-ordered society are moral persons in that, once they have reached the age of reason, each has, and views the others as having, an effective sense of justice, as well as an understanding of a conception of their good. Citizens are equal in that they regard one another as having an equal right to determine, and to assess upon due reflection, the first principles of justice by which the basic structure of their society is to be governed. Finally, the members of a well-ordered society are free in that they think they are entitled to make claims on the design of their common institutions in the name of their own fundamental aims and highest-order interests. At the same time, as free persons, they think of themselves not as inevitably tied to the pursuit of the particular final ends they have

[3] These features were not conveniently stated at any one place in TJ. In this and the next lectures I try to give a clearer and more systematic account of this notion and to indicate its basic role as a model-conception.

at any given time, but rather as capable of revising and changing these ends on reasonable and rational grounds.

There are other features of a well-ordered society, such as its stability with respect to its sense of justice, its existing under the circumstances of justice, and so on. But these matters can be left aside. The essential thing is that, when we formulate the model-conception of the original position, we must view the parties as selecting principles of justice which are to serve as effective public principles of justice in a well-ordered society, and hence for social cooperation among persons who conceive of themselves as free and equal moral persons. Although this description of a well-ordered society is formal, in that its elements taken alone do not imply a specific content for the principles of justice, the description does impose various conditions on how the original position can be set up. In particular, the conception of moral persons as free and equal, and the distinction between rational and full autonomy, must be appropriately reflected in its description. Otherwise the original position cannot fulfill its mediating role to connect a certain conception of the person with definite first principles by means of a procedure in which the parties, as rationally autonomous agents of construction, adopt principles of justice, the public affirmation of which by citizens of a well-ordered society in every-day life enables them to be fully autonomous.

III

Let us descend from these abstractions, at least a bit, and turn to a summary account of the original position. As I have said, justice as fairness begins from the idea that the most appropriate conception of justice for the basic structure of a democratic society is one that its citizens would adopt in a situation that is fair between them and in which they are represented solely as free and equal moral persons. This situation is the original position: we conjecture that the fairness of the circumstances under which agreement is reached transfers to the principles of justice agreed to; since the original position situates free and equal moral persons fairly with respect to one another, any conception of justice they adopt is likewise fair. Thus the name: 'justice as fairness'.

In order to ensure that the original position is fair between individuals regarded solely as free and equal moral persons, we require that, when adopting principles for the basic structure, the parties be deprived of certain information; that is, they are behind what I shall call a "veil of ignorance." For example, they do not know their place in society, their class position, or social status, nor do

they know their fortune in the distribution of natural talents and abilities. It is assumed also that they do not know their conception of the good, that is, their particular final ends; nor finally, their own distinctive psychological dispositions and propensities, and the like. Excluding this information is required if no one is to be advantaged or disadvantaged by natural contingencies or social chance in the adoption of principles. Otherwise the parties would have disparate bargaining advantages that would affect the agreement reached. The original position would represent the parties not solely as free and equal moral persons, but instead as persons also affected by social fortune and natural accident. Thus, these and other limitations on information are necessary to establish fairness between the parties as free and equal moral persons and, therefore, to guarantee that it is as such persons that they agree to society's basic principles of justice.

Now the original position, as described, incorporates pure procedural justice at the highest level. This means that whatever principles the parties select from the list of alternative conceptions presented to them are just. Put another way, the outcome of the original position defines, let us say, the appropriate principles of justice. This contrasts with perfect procedural justice, where there is an independent and already given criterion of what is just (or fair) and where a procedure exists to ensure a result that satisfies this standard. This is illustrated by the familiar example of dividing a cake: if equal division is taken as fair, then we simply require the person who cuts it to have the last piece. (I forego the assumptions necessary to make the example airtight.) The essential feature of pure procedural justice, as opposed to perfect procedural justice, is that there exists no independent criterion of justice; what is just is defined by the outcome of the procedure itself.

One reason for describing the original position as incorporating pure procedural justice is that it enables us to explain how the parties, as the rational agents of construction, are also autonomous (as such agents). For the use of pure procedural justice implies that the principles of justice themselves are to be constructed by a process of deliberation, a process visualized as being carried out by the parties in the original position. The appropriate weight of considerations for and against various principles is given by the force of these considerations for the parties, and the force of all reasons on balance is expressed by the agreement made. Pure procedural justice in the original position allows that in their deliberations the parties are not required to apply, nor are they bound by, any antecedently

given principles of right and justice. Or, put another way, there exists no standpoint external to the parties' own perspective from which they are constrained by prior and independent principles in questions of justice that arise among them as members of one society.

I call your attention to the following: I have said above that there is no standpoint external to the parties' own perspective from which they are bound in questions of justice that arise between them. Here the phrase 'between them' is significant. It signals the fact that I am leaving aside two important matters: questions of justice between societies (the law of nations), and our relations to the order of nature and to other living things. Both these questions are of first importance and immensely difficult; except in a few special cases, no attempt was made in *A Theory of Justice* to discuss these questions.[4] I shall simply proceed on the idea that we may reasonably begin with the basic structure of one society as a closed and self-sufficient system of cooperation. Should we find a suitable conception for this case, we can then work both inward to principles for associations and practices, and outward to the law of nations and order of nature itself. How far this can be done, and to what extent the conception of justice for the basic structure will have to be revised in the process, cannot be foreseen in advance. Here I merely wish to register these limitations of my discussion.

So far the autonomy of the parties is expressed by their being at liberty to agree to any conception of justice available to them as prompted by their rational assessment of which alternative is most likely to advance their interests. In their deliberations they are not required to apply, or to be guided by, any principles of right and justice, but are to decide as principles of rationality dictate, given their situation. But the propriety of the term 'autonomy' as applied to the parties also depends on what their interests are and on the nature of constraints to which they are subject. So let's review these matters.

IV

Recall that the parties are to adopt principles to serve as the effective public conception of justice for a well-ordered society. Now the citizens of such a society regard themselves as moral persons and as having a conception of the good (an ordered scheme of final ends) for the sake of which they think it proper to make claims on the design of their common institutions. So in the original position we

4 See TJ, §58, where several cases of conscientious refusal are considered in connection with the problem of just war. As for our relations with the order of nature, note the last paragraph of §77.

may describe the parties either as the representatives (or trustees) of persons with certain interests or as themselves moved by these interests. It makes no difference either way, although the latter is simpler and I shall usually speak in this vein.

To continue: we take moral persons to be characterized by two moral powers and by two corresponding highest-order interests in realizing and exercising these powers. The first power is the capacity for an effective sense of justice, that is, the capacity to understand, to apply and to act from (and not merely in accordance with) the principles of justice. The second moral power is the capacity to form, to revise, and rationally to pursue a conception of the good. Corresponding to the moral powers, moral persons are said to be moved by two highest-order interests to realize and exercise these powers. By calling these interests "highest-order" interests, I mean that, as the model-conception of a moral person is specified, these interests are supremely regulative as well as effective. This implies that, whenever circumstances are relevant to their fulfillment, these interests govern deliberation and conduct. Since the parties represent moral persons, they are likewise moved by these interests to secure the development and exercise of the moral powers.

In addition, I assume that the parties represent developed moral persons, that is, persons who have, at any given time, a determinate scheme of final ends, a particular conception of the good. Thus the model-conception defines moral persons as also determinate persons, although from the standpoint of the original position, the parties do not know the content of their conception of the good: its final ends. This conception yields a third interest that moves the parties: a higher-order interest in protecting and advancing their conception of the good as best they can, whatever it may be. The reason this is but a higher-order and not a highest-order interest is that, as we shall see later, it is in essential respects subordinate to the highest-order interests.

Now in view of these three regulative interests, the veil of ignorance poses a problem: how are we to set up the original position so that the parties, as representatives of persons with these interests, can make a rational agreement? It is at this point that the account of primary goods is introduced: by stipulating that the parties evaluate conceptions of justice by a preference for these goods, we endow them, as agents of construction, with sufficiently specific desires so that their rational deliberations reach a definite result. We look for social background conditions and general all-purpose means normally necessary for developing and exercising the two moral powers

and for effectively pursuing a conception of the good. Thus a very brief explanation of the parties' preference for the primary goods enumerated in *A Theory of Justice* is this: [5]

(i) The basic liberties (freedom of thought and liberty of conscience, etc.) are the background institutions necessary for the development and exercise of the capacity to decide upon and revise, and rationally to pursue, a conception of the good. Similarly, these liberties allow for the development and exercise of the sense of right and justice under social conditions that are free.

(ii) Freedom of movement and free choice of occupation against a background of diverse opportunities are required for the pursuit of final ends, as well as to give effect to a decision to revise and change them, if one so desires.

(iii) Powers and prerogatives of offices and positions of responsibility are needed to give scope to various self-governing and social capacities of the self.

(iv) Income and wealth, understood broadly as they must be, are all-purpose means (having an exchange value) for achieving directly or indirectly almost any of our ends, whatever they happen to be.

(v) The social bases of self-respect are those aspects of basic institutions which are normally essential if individuals are to have a lively sense of their own worth as moral persons and to be able to realize their higher-order interests and advance their ends with zest and self-confidence.

Granted the correctness of these observations, the parties' preference for primary goods is rational. (I shall assume that in this context our intuitive notion of rationality suffices for our purposes here, and so I shan't discuss it until the next section.)

There are many points about primary goods which need to be examined. Here I mention only the leading idea, namely, that primary goods are singled out by asking which things are generally necessary as social conditions and all-purpose means to enable human beings to realize and exercise their moral powers and to pursue their final ends (assumed to lie within certain limits). Here we must look to social requirements and the normal circumstances of human life in a democratic society. Now note that the conception of moral persons as having certain specified highest-order interests selects what is to count as primary goods within the framework of the

---

[5] A fuller discussion can be found in Allen Buchanan, "Revisability and Rational Choice," *Canadian Journal of Philosophy*, v, 3 (November 1975): 395–408. For a more general account of which the use of primary goods is a special case, see T. M. Scanlon, "Preference and Urgency," this JOURNAL, LXXII, 19 (Nov. 6, 1975): 655–669.

model-conceptions. Thus these goods are not to be understood as general means essential for achieving whatever final ends a comprehensive empirical or historical survey might show people usually or normally to have in common under all social conditions. There may be few if any such ends; and those there are may not serve the purpose of constructing a conception of justice reasonable for us. The list of primary goods does not rest on that kind of general fact, although it does rely on general social facts, once the conception of the person and its highest-order interests are fixed. (Here I should comment that, by making the account of primary goods rest upon a particular conception of the person, I am revising the suggestions in *A Theory of Justice*, since there it can seem as if the list of primary goods is regarded as the outcome of a purely psychological, statistical, or historical inquiry.) [6]

What bearing do these remarks about primary goods have on our original question about rational autonomy? We observed that this autonomy surely depends in part upon the interests that move the parties and not solely on their being bound by no prior and independent principles of right. Were the parties moved solely by lower-order impulses, say for food and drink, or by certain particular affections for this or that group of persons, association, or community, we might think of them as heteronomous and not as autonomous. But at the basis of the desire for primary goods are the highest-order interests of moral personality and the need to secure one's conception of the good (whatever it is). Thus the parties are simply trying to guarantee and to advance the requisite conditions for exercising the powers that characterize them as moral persons. Certainly this motivation is neither heteronomous nor self-centered: we expect and indeed want people to care about their liberties and opportunities in order to realize these powers, and we think they show a lack of self-respect and weakness of character in not doing so. Thus the assumption that the parties are mutually disinterested and, hence, concerned to ensure their own highest-order interests (or those of the persons they represent) should not be confused with egoism.

In conclusion, then, the parties as rational agents of construction are described in the original position as autonomous in two re-

[6] See, for example, §15, pp. 92 ff, where primary goods are first discussed at some length; and also pp. 142 f, 253, 260, and 433 f. The question whether the account of primary goods is a matter for social theory, or depends essentially on a conception of the person, is not discussed. I am grateful to Joshua Cohen, Joshua Rabinowitz, T. M. Scanlon, and Michael Teitelman for helpful criticism and clarification on this important point.

spects: first, in their deliberations they are not required to apply, or to be guided by, any prior and antecedent principles of right and justice. This is expressed by the use of pure procedural justice. Second, they are said to be moved solely by the highest-order interests in their moral powers and by their concern to advance their determinate but unknown final ends. The account of primary goods and its derivation convey this side of autonomy. Given the veil of ignorance, the parties can be prompted only by these highest-order interests, which they must, in turn, render specific by the preference for primary goods.

v

So much for the notion of rational autonomy of the parties as agents of construction. I now turn to the notion of full autonomy; although this notion is realized only by the citizens of a well-ordered society in the course of their daily lives, the essential features of it must nevertheless be represented in a suitable manner in the original position. For it is by affirming the first principles that would be adopted in this situation and by publicly recognizing the way in which they would be agreed to, as well as by acting from these principles as their sense of justice dictates, that citizens' full autonomy is achieved. We must ask, then, how the original position incorporates the requisite elements of full autonomy.

Now these elements are not expressed by how the parties' deliberations and motivation are described. The parties are merely artificial agents, and are presented not as fully but only as rationally autonomous. To explain full autonomy, let us note two elements of any notion of social cooperation. The first is a conception of the *fair terms of cooperation*, that is, terms each participant may reasonably be expected to accept, provided that everyone else likewise accepts them. Fair terms of cooperation articulate an idea of reciprocity and mutuality: all who cooperate must benefit, or share in common burdens, in some appropriate fashion as judged by a suitable benchmark of comparison. This element in social cooperation I call the *Reasonable*. The other element corresponds to the *Rational*: it expresses a conception of each participant's rational advantage, what, as individuals, they are trying to advance. As we have seen, the rational is interpreted by the original position in reference to the desire of persons to realize and to exercise their moral powers and to secure the advancement of their conception of the good. Given a specification of the parties' highest-order interests, they are rational in their deliberations to the extent that sensible principles of rational choice guide their decisions. Familiar

examples of such principles are: the adoption of effective means to ends; the balancing of final ends by their significance for our plan of life as a whole and by the extent to which these ends cohere with and support each other; and finally, the assigning of a greater weight to the more likely consequences; and so on. Although there seems to be no one best interpretation of rationality, the difficulties in explaining Kantian constructivism do not lie here. Thus I ignore these matters, and focus on the more obscure notion of the Reasonable and how it is represented in the original position.

This representing is done essentially by the nature of the constraints within which the parties' deliberations take place and which define their circumstances with respect to one another. The Reasonable is incorporated into the background setup of the original position which frames the discussions of the parties and situates them symmetrically. More specifically, in addition to various familiar formal conditions on first principles, such as generality and universality, ordering and finality, the parties are required to adopt a public conception of justice and must assess its first principles with this condition in mind. (I shall say more about the publicity condition in the next lecture.)

Again, the veil of ignorance implies that persons are represented solely as moral persons and not as persons advantaged or disadvantaged by the contingencies of their social position, the distribution of natural abilities, or by luck and historical accident over the course of their lives. As a result they are situated equally as moral persons, and in this sense fairly. Here I appeal to the idea that, in establishing the truly basic terms of social cooperation, the possession of the minimum adequate powers of moral personality (the powers that equip us to be normally cooperating members of society over a complete life) is the sole relevant characteristic. This presumption, plus the precept that equals in all relevant respects are to be represented equally, ensures that the original position is fair.

The last constraint I shall mention here is this: the stipulation that the first subject of justice is the basic structure of society, that is, the main social institutions and how they cohere together into one system, supports situating the parties equally and restricting their information by the veil of ignorance. For this stipulation requires the parties to assess alternative conceptions as providing first principles of what we may call *background justice:* it is only if the basic structure satisfies the requirements of background justice that a society treats its members as equal moral persons. Otherwise, its fundamental regulative arrangements do not answer to principles its citizens would adopt when fairly represented solely as such persons.

Let us pull together these remarks as follows: the Reasonable presupposes and subordinates the Rational. It defines the fair terms of cooperation acceptable to all within some group of separately identifiable persons, each of whom possesses and can exercise the two moral powers. All have a conception of their good which defines their rational advantage, and everyone has a normally effective sense of justice: a capacity to honor the fair terms of cooperation. The Reasonable presupposes the Rational, because, without conceptions of the good that move members of the group, there is no point to social cooperation nor to notions of right and justice, even though such cooperation realizes values that go beyond what conceptions of the good specify taken alone. The Reasonable subordinates the Rational because its principles limit, and in a Kantian doctrine limit absolutely, the final ends that can be pursued.

Thus, in the original position we view the Reasonable as expressed by the framework of constraints within which the deliberations of the parties (as rationally autonomous agents of construction) take place. Representative of these constraints are the condition of publicity, the veil of ignorance and the symmetry of the parties' situation with respect to one another, and the stipulation that the basic structure is the first subject of justice. Familiar principles of justice are examples of reasonable principles, and familiar principles of rational choice are examples of rational principles. The way the Reasonable is represented in the original position leads to the two principles of justice. These principles are constructed by justice as fairness as the content of the Reasonable for the basic structure of a well-ordered society.

VI

This concludes my account of the distinction between Rational and Full Autonomy and explains how these notions are expressed in the original position. In certain respects, however, the contrast between the Reasonable and the Rational, as drawn in the last two paragraphs, is too stark and may give a misleading impression of how these notions are to be understood. By way of clarification, I consider an objection which parallels the criticism Schopenhauer made against Kant's doctrine of the Categorical Imperative.[7] You will

---

[7] See *On the Basis of Ethics* (1840), Part II, §7, E. F. J. Payne, trans. (New York: Liberal Arts Press, 1965), pp. 89–92. I am indebted to Joshua Cohen for pointing out to me that my previous reply to this criticism misses the force of Schopenhauer's objection. See TJ, pp. 147 f. Thanks to him, I believe the reply in the text is better and connects with the revised account of primary goods. I am indebted also to Stephen Darwall's "A Defense of the Kantian Interpretation," *Ethics*, LXXXVI, 2 (January 1976): 164–170.

recall that Schopenhauer maintained that, in arguing for the duty of mutual aid in situations of distress (the fourth example in the *Grundlegung*), Kant appeals to what rational agents, as finite beings with needs, can consistently will to be universal law. In view of our need for love and sympathy, on at least some occasions, we cannot will a social world in which others are always indifferent to our pleas in such cases. From this Schopenhauer claimed that Kant's view is at bottom egoistic, from which it follows that it is but a disguised form of heteronomy after all.

Here I am concerned not to defend Kant against this criticism but to point out why the parallel objection to justice as fairness is incorrect. To this end, observe that there are, offhand, two things that prompt Schopenhauer's objection. First, he believes that Kant asks us to test maxims in the light of their general consequences for our natural inclinations and needs, when these maxims are made universal laws, and that these inclinations and needs are viewed egoistically. Second, the rules that define the procedure for testing maxims Schopenhauer interprets as external constraints, imposed so to speak from the outside by the limitations of our situation, which we should like to surmount, and not derived from the essential features of ourselves as moral persons. These two considerations lead Schopenhauer to say that the categorical imperative is a principle of reciprocity which egoism cunningly accepts as a compromise; as such a principle, it may be appropriate for a confederation of nation states but not as a moral principle.

Now consider the parallel criticism of justice as fairness in regard to these two points. Concerning the first, though it is indeed true that the parties in the original position are mutually disinterested and evaluate principles of justice in terms of primary goods, they are moved in the first instance by their highest-order interests in developing and exercising their moral powers; and the list of primary goods, and the index of these goods, is to be explained so far as possible by reference to these interests. Since these interests may be taken to specify their needs as moral persons, the parties' aims are not egoistic, but entirely fitting and proper. It accords with the conception of free personality held in a democratic society that citizens should secure the conditions for realizing and exercising their moral powers, as well as the social bases and means of their self-respect. This contrasts with Schopenhauer's presumption that in Kant's doctrine maxims are tested by their consequences for the fulfillment of the agent's natural inclinations and needs.

Turning to the second point, what I have called "the constraints

imposed on the parties in the original position" are indeed external
to the parties as rational agents of construction. Nevertheless, these
constraints express the Reasonable and, therefore, the formal con-
ditions implicit in the moral powers of the members of a well-
ordered society, whom the parties represent. This contrasts with
Schopenhauer's second presumption that the constraints of the cat-
egorical imperative derive from the limitations of our finite nature,
which, prompted by our natural inclinations, we should like to
overcome. In justice as fairness, the Reasonable frames the Rational
and is derived from a conception of moral persons as free and
equal. Once this is understood, the constraints of the original posi-
tion are no longer external. Thus neither basis for Schopenhauer's
objection applies.

Finally, the way in which the Reasonable frames the Rational in
the original position represents a feature of the unity of practical
reason. In Kant's terms, empirical practical reason is represented by
the rational deliberations of the parties; pure practical reason is
represented by the constraints within which these deliberations take
place. The unity of practical reason is expressed by defining the
Reasonable to frame the Rational and to subordinate it absolutely;
that is, the principles of justice that are agreed to are lexically prior
in their application in a well-ordered society to claims of the good.
This means, among other things, that the principles of justice and
the rights and liberties they define cannot, in such a society, be over-
ridden by considerations of efficiency and a greater net balance of
social well-being. This illustrates one feature of the unity of reason:
the Reasonable and the Rational are unified within one scheme of
practical reasoning which establishes the strict priority of the Rea-
sonable with respect to the Rational. This priority of the right over
the good is characteristic of Kantian constructivism.

Now in a well-ordered society we stipulate that the justification
of the principles of justice as the outcome of the original position
is publicly understood. So not only do citizens have a highest-order
desire, their sense of justice, to act from the principles of justice,
but they understand these principles as issuing from a construction
in which their conception of themselves as free and equal moral
persons who are both reasonable and rational is adequately repre-
sented. By acting from these principles, and affirming them in public
life, as so derived, they express their full autonomy. The rational
autonomy of the parties is merely that of artificial agents who in-
habit a construction designed to model this more inclusive concep-
tion. It is the inclusive conception which expresses the ideal to be
realized in our social world.

It is natural to reply that, all the same, fully autonomous citizens in a well-ordered society act from some desire, and so are still heteronomous, since they are not moved by reason alone.[8] To this the answer is that a Kantian view does not deny that we act from some desire. What is of moment is the kinds of desires from which we act and how they are ordered; that is, how these desires originate within and are related to the self, and the way their structure and priority are determined by principles of justice connected with the conception of the person we affirm. The mediating conception of the original position enables us to connect certain definite principles of justice with a certain conception of free and equal moral persons. Given this connection, an effective sense of justice, the desire to act from the principles of justice, is not a desire on the same footing with natural inclinations; it is an executive and regulative highest-order desire to act from certain principles of justice in view of their connection with a conception of the person as free and equal. And that desire is not heteronomous: for whether a desire is heteronomous is settled by its mode of origin and role within the self and by what it is a desire for. In this case the desire is to be a certain kind of person specified by the conception of fully autonomous citizens of a well-ordered society.

VII

I conclude with a few observations which may help to keep in focus the discussion so far. First, it is important to distingush three points of view: that of the parties in the original position, that of citizens in a well-ordered society, and finally, that of ourselves—you and me who are examining justice as fairness as a basis for a conception of justice that may yield a suitable understanding of freedom and equality.

The first two points of view occur within the doctrine of justice as parts of two of its model-conceptions. Whereas the conceptions of a well-ordered society and of moral persons are fundamental, the original position is the mediating conception once we stipulate that the parties as rational agents of construction are subject to reasonable constraints and are to view themselves as adopting principles to serve as the public conception of justice for a well-ordered society. The intent of justice as fairness is badly misunderstood if the deliberations of the parties and their rational autonomy are confused with full autonomy. Full autonomy is a moral ideal and part of the more comprehensive ideal of a well-ordered society. Rational autonomy is not, as such, an ideal at all, but a device of representa-

---

8 This seems to be the view of Oliver A. Johnson in his reply to Darwall, see fn 7 above. See *Ethics*, LXXXVII, 3 (April 1977): 251–259, p. 253 f.

tion used to connect the conception of the person with definite principles of justice. (Of course, this is not to deny that rational deliberation, suitably circumscribed, is an aspect of the ideal of full autonomy.)

The third point of view—that of you and me—is that from which justice as fairness, and indeed any other doctrine, is to be assessed. Here the test is that of general and wide reflective equilibrium, that is, how well the view as a whole meshes with and articulates our more firm considered convictions, at all levels of generality, after due examination, once all adjustments and revisions that seem compelling have been made. A doctrine that meets this criterion is the doctrine that, so far as we can now ascertain, is the most reasonable for us.

A final observation: it is also useful to distinguish between the roles of a conception of the person and of a theory of human nature.[9] In justice as fairness these ideas are distinct elements and enter at different places. For one thing, the conception of the person is a companion moral ideal paired with that of a well-ordered society. Like any other ideal, it must be possible for people to honor it sufficiently closely; and hence the feasible ideals of the person are limited by the capacities of human nature and the requirements of social life. To this extent such an ideal presupposes a theory of human nature, and social theory generally, but the task of a moral doctrine is to specify an appropriate conception of the person that general facts about human nature and society allow. Starting from the assumption that full autonomy is a feasible ideal for political life, we represent its various aspects in the original position under the headings of the Reasonable and the Rational. Thus this ideal is mirrored in how this position is set up.

A theory of human nature, by contrast, appears in the general facts available to the parties for them to use in assessing the consequences of the various principles of justice and so in deciding which principles are best able to secure their highest-order interests and to lead to a well-ordered society that is stable with respect to its public conception of justice. When we formulate justice as fairness from the third point of view, we supply the parties with the requisite general facts that we take to be true, or true enough, given the state of public knowledge in our society. The agreement of the parties is relative, then, to these beliefs. There is no other way to proceed, since we must start from where we are. But, leaving this

[9] I am indebted to Norman Daniels for clarification of this point.

aside, the point is that a theory of human nature is not part of the framework of the original position, except as such theories limit the feasibility of the ideals of person and society embedded in that framework. Rather, a theory of human nature is an element to be filled in, depending upon the general facts about human beings and the workings of society which we allow to the parties in their deliberations.

In this lecture I have focused on the distinction between rational and full autonomy and have said very little about the notions of the freedom and equality of persons, and even less about how these notions are represented in the original position. These matters I consider in the next lecture.

REPRESENTATION OF FREEDOM AND EQUALITY

IN the last lecture, I focused largely on the distinction between rational and full autonomy. Rational autonomy is expressed in the deliberations of the parties as artificial agents of construction within the original position. Full autonomy is the more comprehensive notion and expresses an ideal of the person affirmed by the citizens of a well-ordered society in their social life. But although I described the parties as representatives of free and equal moral persons, I indicated only briefly what freedom and equality mean and how these features of the person are represented in the original position. Nor did I say very much about the formal condition of publicity, which is a distinctive element of a Kantian view. Exploring these matters will help to fill in the account of the original position and show how justice as fairness is an illustration of Kantian constructivism in moral theory.

I

I shall begin with some further remarks about the model-conception of a well-ordered society. You will recall that I said last time that there are various forms of constructivism. A number of views not usually thought of as constructivist can be presented in this way.[1] This suggests that the three main model-conceptions of justice as fairness—those of a well-ordered society, the conception of the person, and the original position—are all special renderings of more general notions. What characterizes a Kantian doctrine is the particular way in which it interprets these three model-conceptions; especially characteristic, of course, is its conception of the person

[1] Thus, for example, average utilitarianism might be presented as a kind of constructivism. See TJ, §27.

as reasonable and rational, and fully autonomous. I shall not ex-
amine here what these more general model-notions are or how they
might be defined; I mention these questions only to remind us that
the model-conceptions I discuss are special cases that define a par-
ticular moral doctrine.

To continue: recall that a well-ordered society is conceived as an
on-going society, a self-sufficient association of human beings which,
like a nation-state, controls a connected territory. Its members view
their common polity as extending backward and forward in time
over generations, and they strive to reproduce themselves, and their
cultural and social life in perpetuity, practically speaking; that is,
they would envisage any final date at which they were to wind up
their affairs as inadmissible and foreign to their conception of their
association. Finally, a well-ordered society is a closed system; there
are no significant relations to other societies, and no one enters
from without, for all are born into it to lead a complete life.

Next, we assume that, as an on-going society, the scheme of social
and economic activities set up and framed by the basic structure is
productive and fruitful. This implies, for example, that a well-
ordered society does not have a manna economy, nor are its eco-
nomic arrangements a zero-sum game in which none can gain un-
less others lose. Yet it does exist under circumstances of justice, of
which there are two kinds: first, the objective circumstances of mod-
erate scarcity; and, second, the subjective circumstances, namely,
that persons and associations have contrary conceptions of the good
as well as of how to realize them, and these differences set them at
odds, and lead them to make conflicting claims on their institutions.
They hold opposing religious and philosophical beliefs, and affirm
not only diverse moral and political doctrines, but also conflicting
ways of evaluating arguments and evidence when they try to recon-
cile these oppositions. In view of the circumstances of justice, the
members of a well-ordered society are not indifferent to how the
fruits of their social cooperation are distributed, and, for their
society to be stable, the distribution that results and is expected in
the future must be seen to be (sufficiently) just.

Thus, as we noted last time, the stability of a well-ordered society
is not founded merely on a perceived balance of social forces the
upshot of which all accept since none can do better for themselves.
To the contrary, citizens affirm their existing institutions in part
because they reasonably believe them to satisfy their public and
effective conception of justice. Now the notion of publicity has

three levels, which may be distinguished as follows: [2]

The first was mentioned last time: it means that society is effectively regulated by public principles of justice; that is, everyone accepts and knows that the others likewise accept the same principles, and this knowledge in turn is publicly recognized. Also, the institutions that constitute the basic structure of society actually satisfy these principles of justice, and everyone with reason acknowledges this on the basis of commonly shared beliefs confirmed by methods of inquiry and ways of reasoning agreed to be appropriate for questions of social justice.

The second level of publicity concerns the general beliefs in the light of which first principles of justice themselves can be accepted, that is, the theory of human nature and of social institutions generally. Citizens in a well-ordered society roughly agree on these beliefs because they can be supported (as at the first level) by publicly shared methods of inquiry and ways of reasoning thought to be appropriate for this case. These methods and ways of reasoning I assume to be familiar from common sense and to include the procedures and conclusions of science, when these are well established and not controversial. Keep in mind that we aim to find a conception of justice for a democratic society under modern conditions; so we may properly assume that in its public culture the methods and conclusions of science play an influential role. It is precisely these general beliefs, which reflect the current public views in a well-ordered society, that we allow to the parties in the original position for the purpose of assessing alternative principles of justice.

The third and last level of publicity has to do with the complete justification of the public conception of justice as it would be presented in its own terms. This justification includes everything that we would say—you and me—when we set up justice as fairness, and reflect why we do this one way rather than another. At the third level I suppose this full justification also to be publicly known or, better, at least publicly available; this weaker condition allows for the possibility that some will not want to carry moral reflection so far, and certainly they are not required to do so. But if they wish to, the justification is present in public culture, reflected in law and political institutions, and in the philosophical and historical traditions of their interpretation. More specifically, the full justification includes connecting the moral doctrine's model-conceptions with the society's particular conception of the person and of social co-

[2] I am indebted to Joshua Rabinowitz for clarification concerning these distinctions.

operation. This conception is shown in how citizens think of themselves as members of a democratic polity when they examine the doctrine as a whole and find after due reflection that it matches their considered judgments at all levels of generality.

A well-ordered society satisfies what I shall call the *full publicity condition* when all three levels are exemplified. (I reserve the adjective 'full' for the elements of the complete and full rendering of the conception of a well-ordered society.) Now this full condition may seem excessively strong; so let's ask why it is adopted. One reason is that the model-conception of a well-ordered society is to incorporate various formal moral notions into an ideal of social cooperation between persons regarded in a certain way. This ideal is to hold for free and equal moral persons, and views social cooperation not simply as productive and socially coordinated activity, but as fulfilling a notion of fair terms of cooperation and of mutual advantage, as expressed by the distinction between the Reasonable and the Rational. So we should like to find a conception of justice that answers to the full condition; it seems bound to define more specific constraints on conceptions of justice and, hence, is more likely to provide a sharper basis for deciding among conflicting understandings of freedom and equality and for determining how their claims are to be balanced against one another. Recall that this conflict of understandings sets the present practical task of political philosophy.

Another reason for the full publicity condition (and indeed for any of its levels) is that is seems particularly appropriate for a conception of political and social justice. No doubt, publicity is less compelling for other moral notions. But the principles of justice apply to the political constitution and the basic institutions of society, which normally include, even under favorable conditions, some machinery of legal coercion, if only to guarantee the stability of social cooperation.[3] Moreover, these institutions can have decisive long-term social effects and importantly shape the character and aims of the members of society, the kinds of persons they are and want to be. It seems fitting, then, that the fundamental terms of social cooperation between free and equal moral persons should answer to the requirements of full publicity. For if institutions rely

---

[3] Here I should explain that in a well-ordered society coercive sanctions are rather rarely, if ever, actually applied (since offences are presumably infrequent), nor need severe sanctions be legally permitted. Stability means that institutional rules are generally complied with, and the role of the machinery of sanctions is to support citizens' mutual expectations of one another's settled intention to follow these norms. See TJ, pp. 269 f, 336, 576 f.

on coercive sanctions, however seldom necessary and however scrupulously applied, and influence people's deepest aspirations, the grounds and tendency of these institutions should stand up to public scrutiny. When political principles satisfy the full publicity condition, and social arrangements and individual actions are similarly justifiable, the citizens can fully account for their beliefs and conduct to everyone else with assurance that this avowed reckoning itself will strengthen and not weaken the public understanding. The maintenance of the social order does not depend on historically accidental or institutionalized delusions, or other mistaken beliefs about how its institutions work. Publicity ensures, so far as the feasible design of institutions can allow, that free and equal persons are in a position to know and to accept the background social influences that shape their conception of themselves as persons, as well as their character and conception of their good. Being in this position is a precondition of freedom; it means that nothing is or need be hidden.[4]

Thus, given the circumstances of justice, the full publicity condition applies only to the principles of political and social justice and not to all moral notions. Now, although moderate scarcity may possibly be overcome or largely mitigated, justice as fairness assumes that deep and pervasive differences of religious, philosophical, and ethical doctrine remain. For many philosophical and moral notions public agreement cannot be reached; the consensus to which publicity applies is limited in scope to the public moral constitution and the fundamental terms of social cooperation. That citizens in a well-ordered society can agree before one another on principles of justice and recognize their institutions to be just means that they have also agreed that, for certain parts of their common life, considerations of justice are to have a special place. Other reasons are taken not to be appropriate, although elsewhere they may have a governing role, say, within the life of associations. In public questions, ways of reasoning and rules of evidence for reaching true general beliefs that help settle whether institutions are just should be of a kind that everyone can recognize. Although, in a democratic society under modern conditions, these norms are the shared principles and practices of common sense and science (when not controversial), to apply them to other convictions is a different matter.

To conclude: the conception of a well-ordered society includes

[4] Put in a different way: a well-ordered society does not require an ideology in order to achieve stability, understanding 'ideology' (in Marx's sense) as some form of false consciousness or delusory scheme of public beliefs.

and generalizes the idea of religious liberty; it assigns to people's
conception of the good a public status analogous to that of religion.
Although a well-ordered society is divided and pluralistic, its cit-
izens have nevertheless reached an understanding on principles to
regulate their basic institutions. While they cannot achieve agree-
ment in all things, the public agreement on questions of political
and social justice supports ties of civic friendship and secures the
bonds of association.

II

I now consider how the publicity condition is represented in the
original position and examine some queries by way of clarification.
Actually, the representation of publicity (at any level) is quite
straightforward: we simply require the parties as agents of con-
struction to assess conceptions of justice, subject to the constraint
that the principles they agree to must serve as a public conception
of justice in the stipulated sense. Principles which might work quite
well provided they were not publicly acknowledged (as defined at
the first level) or provided the general beliefs upon which they are
founded are not commonly understood, or which would be recog-
nized as fallacious (as defined by the second level) are to be rejected.
Thus the parties must evaluate the social and psychological conse-
quences of various kinds of public knowledge against a certain back-
ground of common beliefs, and these consequences will affect which
conception of justice they adopt, all things considered.

    Since the representation of publicity seems simple enough, it is
more instructive to take up a few points that naturally arise. To
begin with, even the first level of publicity cannot be satisfied in
society unless the parties also agree upon rules of evidence and
forms of reasoning to be used in deciding whether existing institu-
tions fulfill the principles of justice. An agreement on a conception
of justice is fruitless in the absence of an understanding about the
application of its principles. Now, given the subjective circum-
stances of justice (the existence of deep and pervasive religious and
philosophical differences, and the like), the admissible grounds for
holding institutions just or unjust must be limited to those allowed
by forms of reasoning accepted by common sense, including the
procedures of science when generally accepted. Otherwise no effec-
tive undertaking has been made. In a well-ordered society citizens'
judgments of their basic institutions in questions of justice rest on
common knowledge and on shared practices of inquiry. As I have
noted, these restrictions apply only to political and social justice.
On philosophical or religious or other grounds people may, of

course, think certain institutions and policies wrong, but when their beliefs are not commonly based (in the sense defined), they refrain from urging these considerations. The claims of justice have priority and are accepted as decisive in questions concerning the design of the basic structure. The parties recognize, then, that the agreement in the original position has two parts: first, an agreement on principles of justice and, second, a companion agreement on ways of reasoning and rules for weighing evidence which govern the application of those principles. The subjective circumstances of justice limit this companion agreement to the shared beliefs and the recognized procedures of science and common sense.

These remarks are connected with the restrictions contained in the veil of ignorance, as follows. The second level of full publicity is that the general beliefs of social theory and moral psychology relied on by the parties in order to rank conceptions of justice must also be publicly known. Citizens in a well-ordered society know what beliefs are thought to support the recognized principles of justice and belong, therefore, to their complete public justification. This presupposes that, when the original position is set up, we stipulate that the parties must reason only from general beliefs that are suitably common. Thus, the question arises: what is the reason for limiting the parties to these beliefs and not allowing them to take into account all true beliefs? Surely some religious and philosophical doctrines must be true, even if they merely deny other false or incoherent doctrines. Why isn't the most reasonable conception that which is founded on the whole truth and not simply on a part of it, much less on merely commonly based beliefs that happen to be publicly accepted at any given time, for these presumably contain at least some error?

A fully adequate answer to this question involves a number of matters that I cannot go into here. Therefore I restrict my reply to the practical answer implicit in what has already been said.[5] In view of the practical task of political philosophy, it would be a mistake to dismiss this answer as merely practical. But to proceed: as I shall note in the last lecture, Kantian constructivism allows us to say that the (or a) most reasonable conception of justice (should one or more exist) is the conception that the parties would adopt were they to know all the relevant and true beliefs concerning human nature and social theory. This conception of justice has a natural pre-eminence. It is essential to see, however, that not even this concep-

[5] I am indebted to Thomas Nagel, Derek Parfit, and T. M. Scanlon for instructive discussion on this and related points.

tion is accepted on the basis of the whole truth, if the whole truth is to include the truths of religion and philosophy and of moral and political doctrine. By assumption, in a well-ordered democratic society under modern conditions, there is no settled and enduring agreement on these matters; this stipulation is contained in the subjective circumstances of justice. If we ask why these circumstances are assumed, the reply is that, unlike the objective circumstances of moderate scarcity, which might be overcome, the subjective circumstances seem bound to obtain in the absence of a sustained and coercive use of state power that aims to enforce the requisite unanimity. There is no alternative, then, to founding a conception of justice suitable for a well-ordered democratic society on but a part of the truth, and not the whole, or, more specifically, on our present commonly based and shared beliefs, as above defined.

It is important to observe that this practical answer does not imply either skepticism or indifference about religious, philosophical, or moral doctrines. We do not say that they are all doubtful or false, or address questions to which truth and falsehood do not apply. Instead, long historical experience suggests, and many plausible reflections confirm, that on such doctrines reasoned and uncoerced agreement is not to be expected. Religious and philosophical views express outlooks toward the world and our life with one another, severally and collectively, as a whole. Our individual and associative points of view, intellectual affinities and affective attachments, are too diverse, especially in a free democratic society, to allow of lasting and reasoned agreement. Many conceptions of the world can plausibly be constructed from different standpoints. Diversity naturally arises from our limited powers and distinct perspectives; it is unrealistic to suppose that all our differences are rooted solely in ignorance and perversity, or else in the rivalries that result from scarcity. Justice as fairness tries to construct a conception of justice that takes deep and unresolvable differences on matters of fundamental significance as a permanent condition of human life. Indeed, this condition may have its good side, if only we can delineate the character of social arrangements that enable us to appreciate its possible benefits.

One final comment: in order to explain why the veil of ignorance excludes certain kinds of beliefs, even when we as individuals are convinced they are true, I have cited the public role that a conception of justice has in a well-ordered society. Because its principles are to serve as a shared point of view among citizens with opposing religions, philosophical and moral convictions, as well as diverse

conceptions of the good, this point of view needs to be appropriately impartial among those differences. Now this brings out in striking fashion the practical purposes and social role that a conception of social justice must fulfill. The very content of the first principles of justice, in contrast with the content of derivative standards and precepts, is determined in part by the practical task of political philosophy. We are accustomed to the idea that secondary norms and working criteria, by which our moral views are applied, must be adjusted to the normal requirements of social life as well as to the limited capacities of human reasoning, and the like. But we tend to regard these adjustments as made in the light of various first principles, or a single such principle. First principles themselves are not widely regarded as affected by practical limitations and social requirements. In Kantian constructivism at least, the situation is different, as we shall see next time: the first principles of justice are thought to depend on such practical considerations.

III

Let us now turn to freedom and equality. I have said that citizens in a well-ordered society regard themselves as free and equal moral persons. Last time we took up the notion of moral persons as characterized by two moral powers: the capacity to act from a sense of justice, and a capacity to form and rationally to pursue a conception of the good. Moral persons are moved by two corresponding highest-order interests to develop and to exercise these powers. We surveyed how moral personality is represented in the original position by elements falling under the Reasonable and the Rational and how in turn this distinction connects with the contrast between Rational and Full Autonomy.

I begin with freedom: I said that citizens in a well-ordered society view themselves as free in two ways. First of all, they hold themselves entitled to make claims on the design of social institutions in the name of their highest-order interests and final ends, when these ends lie within certain limits. We can elaborate this by saying: citizens think of themselves as self-originating sources of valid claims. Provided their final ends are not things directly contrary to the public principles of justice, these ends along with their highest-order interests support such claims, the weight of which may depend, of course, on particular circumstances. People are self-originating sources of claims in the sense that their claims carry weight on their own without being derived from prior duties or obligations owed to society or to other persons, or, finally, as derived from, or assigned to, their particular social role. Claims that are said to be founded

on duties to self, if some hold that there are such duties, are counted as self-originating for the purposes of a conception of social justice.

Thus, one aspect of freedom is that of the person as a self-originating source of claims. We can see this by contrasting this basis of claims with one derived from our social role, for example, that of claims implied by the duties we must discharge in certain positions of authority, or of those which result from obligations we have assumed. Again, people who act as agents for others have rights and powers dependent upon the rights and intentions of those who have authorized them as their agents. To take the extreme case, slaves are human beings who are not counted as self-originating sources of claims at all; any such claims originate with their owners or in the rights of a certain class in society. Of course, this extreme condition is usually mitigated to some degree, but even when the legal system allows slaves to originate claims, the explanation may rest not on claims that slaves have as moral persons but on the recognition of the unhappy consequences for the rest of society of an extreme institution of slavery. The contrast with slavery makes clear why counting moral personality itself as a source of claims is an aspect of freedom.

A second aspect of freedom, as I described it last time, is that, as free persons, citizens recognize one another as having the moral power to have a conception of the good. This means that they do not view themselves as inevitably tied to the pursuit of the particular conception of the good and its final ends which they espouse at any given time.[6] Instead, as citizens, they are regarded as, in general, capable of revising and changing this conception on reasonable and rational grounds. Thus it is held to be permissible for citizens to stand apart from conceptions of the good and to survey and assess their various final ends; indeed this must be done whenever these ends conflict with the principles of justice, for in that case they must be revised. And here I should explain that by a conception of the good is meant not merely a system of final ends but also a view about one's relation to others and to the world which makes these ends appropriate.

In sum, then, citizens as free persons have the right to view their persons as independent and not identified with any particular system of ends. Given their moral power to form, to revise, and rationally to pursue a conception of the good, their public identity

as a moral person and a self-originating source of claims is not affected by changes over time in their conceptions of the good, at least so long as these changes are in certain ways continuous and have suitable explanations. These remarks are unhappily extremely vague; their only purpose, however, is to indicate the conception of the person connected with the public conception of justice in a well-ordered society, and so with the principles of justice that apply to its basic institutions. By contrast, citizens in their personal affairs, or within the internal life of associations, may regard their ends and aspirations differently. They may have attachments and loves that they believe they would not, or could not, stand apart from; and they might regard it as unthinkable for them to view themselves without certain religious and philosophical convictions and com- mitments. But none of this need affect the conception of the person connected with society's public conception of justice and its ideal of social cooperation. Within different contexts we can assume di- verse points of view toward our person without contradiction so long as these points of view cohere together when circumstances require. As always, our focus here is on the public conception that underlies the principles of social justice.[7]

A third aspect of freedom I shall only mention here: namely, as responsibility for ends. Very roughly, this means that, given just background institutions and the provision for all of a fair index of primary goods (as required by the principles of justice), citizens are capable of adjusting their aims and ambitions in the light of what they can reasonably expect and of restricting their claims in matters of justice to certain kinds of things. They recognize that the weight of their claims is not given by the strength or intensity of their wants and desires, even when these are rational. But to explain these matters here would take us too far afield.[8] I shall consider only two aspects of freedom: one as self-originating source of claims, the other as independence.

I have yet to state the sense in which citizens in a well-ordered society are equal moral persons. But before I do this, let's introduce a further idealization of the notion of a well-ordered society. Our aim is to ascertain the conception of justice most appropriate for a

[7] The remarks in this paragraph indicate part of the basis for a reply that I believe can be made to some of the objections raised by Bernard Williams to a Kantian view. See his paper "Persons, Character and Morality," in A. O. Rorty, ed., The Identities of Persons (Berkeley: University of California Press, 1976), pp. 197–216.

[8] For a brief account, see "Fairness to Goodness," Philosophical Review, LXXXIV, 4 (October 1975): 536–554, pp. 551–554.

democratic society in which citizens conceive of themselves in a cer-
tain way. So let's add that all citizens are fully cooperating members
of society over the course of a complete life. This means that every-
one has sufficient intellectual powers to play a normal part in soci-
ety, and no one suffers from unusual needs that are especially diffi-
cult to fulfill, for example, unusual and costly medical requirements.
Of course, care for those with such requirements is a pressing prac-
tical question. But at this initial stage. the fundamental problem of
social justice arises between those who are full and active and
morally conscientious participants in society, and directly or indi-
rectly associated together throughout a complete life. Therefore, it
is sensible to lay aside certain difficult complications. If we can work
out a theory that covers the fundamental case, we can try to extend
it to other cases later. Plainly a theory that fails for the fundamen-
tal case is of no use at all.

To return to equality, we say: everyone is equally capable of
understanding and complying with the public conception of jus-
tice; therefore all are capable of honoring the principles of justice
and of being full participants in social cooperation throughout
their lives. On this basis, together with each person's being a self-
originating source of valid claims, all view themselves as equally
worthy of being represented in any procedure that is to determine
the principles of justice that are to regulate the basic institutions
of their society. This conception of equal worth is founded on the
equally sufficient capacity (which I assume to be realized) to under-
stand and to act from the public conception of social cooperation.

Now some citizens have a deeper understanding of justice than
others, and a greater facility in applying its principles and making
reasonable decisions, especially in hard cases. The judicial virtues
depend upon special gifts and acquired wisdom. Equality means
that, although these virtues may render some better qualified than
others for certain more demanding offices and positions (those of a
judicial kind, for example), nevertheless, given people's actual place
in just institutions, including the status all have as equal citizens,
everyone's sense of justice is equally sufficient relative to what is
asked of them. This suffices for everyone to be equally worthy of
representation in a procedure that is to settle the fundamental
terms of social cooperation, given that all are able to be fully co-
operating members of society over a complete life.

Finally, citizens in a well-ordered society are in their conduct
(more or less) above reproach. Whatever their actions, all conform
to the acknowledged requirements of justice for the most part. This

follows from the assumption that everyone has an equally effective sense of justice. The usual differences in the degree to which people are open to censure in matters of justice do not obtain. Nevertheless, certain social and economic inequalities presumably exist, but, whatever their explanation, they do not match differences in the degree to which people comply with just arrangements. Since justice regulates these inequalities, the public conception, whatever it is, cannot read: to persons according to their moral worth. This much follows from the general description of a well-ordered society.

IV

It remains to consider the representation of freedom and equality in the original position. I observe first, however, that the two powers of moral persons are represented in a purely formal way. Thus while the parties as agents of construction are assumed to have an effective sense of justice, this is taken to mean that they have a capacity to understand and apply the various principles of justice that are under discussion, as well as a sufficiently strong desire to act upon whatever principles are eventually adopted. Since these principles are not yet agreed to, the parties' sense of justice lacks content. Their formal sense of justice simply ensures that, as members of society, they can follow the most reasonable conception of justice, everything taken into account. The original agreement meets this condition on a bona fide undertaking.

The second capacity of moral personality is likewise represented in a formal fashion. Although the parties have the power to develop, revise, and pursue rationally a conception of the good, they do not know its particular final ends. The capacity for such a conception is assumed to be realized in society, and indeed to have some determinate content. These restrictions on information, which are consequences of the veil of ignorance, require us to characterize the parties' moral powers in a formal way.

To prevent misunderstanding, I reiterate what I said last time: that the motivation of the parties is appropriate to the representation of moral persons. Once such persons are characterized by the moral powers, it is proper that they should strive to realize and exercise these capacities, and be moved by what I have called their "highest-order" interests. This leads us to say that the parties are mutually disinterested, that is, that they aim to secure the interests of their moral personality and to try to guarantee the objective social conditions that enable them rationally to assess their final ends and to do their part in cooperating with others in fair social arrangements to produce the all-purpose means to achieve them. Since

the parties are determinate persons, they also try to ensure their own ability to pursue their particular aims and to protect the objects of their affections, whatever these are. Given the limits on information, they settle an index of primary goods as the most effective way of achieving these objectives.

Now the freedom of persons as self-originating sources of claims is represented by not requiring the parties to justify the claims they wish to make. Whether they are citizens acting as deputies for themselves or whether they are trustees, they are free to act in the best interests of whomever they represent within the framework of reasonable constraints embedded in the original position. It belongs to the parties' rational autonomy that there are no given antecedent principles external to their point of view to which they are bound. The interests they try to advance need not be derived from some prior duty or obligation, either to other persons or to society. Nor do the parties recognize certain intrinsic values as known by rational intuition, for example, the perfectionist values of human excellence or of truth and beauty. This is how freedom as originating claims is represented. Although some or all in society may recognize these values, their acceptance is, from the standpoint of political and social justice, self-imposed, or else a consequence of the principles of justice still to be adopted.

Freedom as independence is represented in how the parties are moved to give priority to guaranteeing the social conditions for realizing their highest-order interests, and in their having grounds for agreement despite the severe restrictions on information implied by the veil of ignorance. To explain this, consider the objection that, since these restrictions exclude knowledge of final ends, no rational agreement is possible. The reply to this objection is that it ties the aspirations of the person too closely to the particular conception of the good that is being pursued at any given time. As free persons have been characterized, rational deliberation is still possible even when the final ends of this conception are unknown. The explanation is that free persons have a regulative and effective desire to be a certain kind of person, so that the veil of ignorance does not eliminate all bases for deliberation. For if it did, the parties would lack the highest-order interests in guaranteeing the objective social conditions for developing and exercising their moral powers and in securing the normally essential all-purpose means for advancing their plan of life.

In a Kantian constructivist view, then, it is a feature attributed to persons (for the purposes of a conception of social justice) that they

can stand above and critically survey their own final ends by reference to a notion of the Reasonable and the Rational. In this sense, they are independent from and moved by considerations other than those given by their particular conception of the good. The veil of ignorance forces the parties to do something analogous, but on a more abstract level: since they are ignorant of their final ends and of much else, they must try to work out which conception of justice is most likely to secure the social conditions and all-purpose means necessary to realize their highest-order interests and determinate but unknown conception of the good.

A further feature of a Kantian doctrine is that it aims at the thickest possible veil of ignorance.[9] This may be explained as follows: there are two distinct rationales for excluding information, and one leads to a thicker veil of ignorance than the other. The rationale drawn from Hume's "judicious spectator" is designed to prevent the parties from reasoning according to the principle: to persons according to their threat advantage. By denying everyone a knowledge of these contingencies, a kind of impartiality is achieved. We begin by allowing the parties all information about themselves: their social position, realized natural assets, their ends and aims, and so on. Enough information is then ruled out to achieve impartiality in the sense of the elimination of threat advantage. The veil of ignorance is thin, because no more knowledge is excluded than is necessary to secure this result; the parties still know the general configuration of society, its political structure and economic organization, and so on. So long as the relevant particular facts are unknown, the influence of threat advantage is eliminated.

The Kantian rationale proceeds in the opposite direction: it starts by allowing the parties no information and then adds just enough so that they can make a rational agreement. The first principles of justice should be those of rationally autonomous agents moved to secure the conditions for the development and exercise of their moral powers, and their determinate (but unknown) final ends. It does not suffice that they are impartial in the sense of being unable to take advantage of their superior position (if such they have). The parties are not to be influenced by any particular information that is not part of their representation as free and equal moral persons with a determinate (but unknown) conception of the good, unless this information is necessary for a rational agreement to be reached.

And so the veil of ignorance is presumably thicker: the Humean ra-
tionale will not exclude certain particular information; the Kantian
will not include it. Even if these different restrictions led to the
same principles, the thicker veil of ignorance would still be pref-
erable, since these principles are then connected more clearly to the
conception of free and equal moral persons. Were we to allow
knowledge of the general institutional features of society, we would
permit particular information about the outcome of society's his-
tory to obscure how intimately the principles adopted are tied to
the conception of the person. And should this information yield
different principles, it would do so on inappropriate grounds. In
either case it should be excluded in order to have a lucid represen-
tation of the notion of freedom that characterizes a Kantian view.

                                V

The representation of equality is an easy matter: we simply describe
all the parties in the same way and situate them equally, that is,
symmetrically with respect to one another. Everyone has the same
rights and powers in the procedure for reaching agreement. Now it
is essential to justice as fairness that the original position be fair
between equal moral persons so that this fairness can transfer to the
principles adopted. Let's recall, then, why the original position is
said to be fair.

To begin with, we take the basic structure of society as the first
subject of justice. Next, we say that to determine first principles for
this subject, the only relevant feature of human beings is their hav-
ing the minimum sufficient capacity for moral personality (as ex-
pressed by the two moral powers), given that, as I suppose, all are
fully capable of being fully cooperating members of society over a
complete life. Finally, we assume that persons equal in all relevant
respects are to be represented equally. These presumptions ensure
that the original position is fair between equal moral persons and,
therefore, that it correctly represents how the members of a well-
ordered society regard one another. Doubts about the fairness of the
original position are perhaps best dealt with by defense against
various objections.

For example, it is sometimes said that the original position is un-
fair to those with superior natural endowments, since, by excluding
knowledge of such gifts, it precludes them from affecting the outcome.
Again, justice as fairness is said to be unfair to those who have
conscientiously acquired certain skills in the expectation of benefit-
ing from them. But these objections fail, I think, to allow for the
special features of the problem of background justice. Keep in mind

that we seek principles to regulate the basic structure into which we are born to lead a complete life. The thesis is that the only relevant feature in connection with these principles is the capacity for moral personality in the sense defined. The way in which we think about fairness in everyday life ill prepares us for the great shift in perspective required for considering the justice of the basic structure itself.

Once this is understood, we must distinguish between features of persons relevant for the justice of the basic structure and features relevant for the fairness of the actual distributions of benefits that come about within this structure as a result of the particular decisions and activities of individuals and associations. These distributions arise from the honoring of legitimate expectations and are, of course, affected by what people actually decide to do, given their knowledge of existing institutional rules, as well as by the various realized skills and talents of individuals. A further essential distinction is between the unequal distribution of natural assets, which is simply a natural fact and neither just nor unjust, and the way the basic structure of society makes use of these natural differences and permits them to affect the social fortune of citizens, their opportunities in life, and the actual terms of cooperation between them. Plainly it is the way that social institutions use natural differences, and allow accident and chance to operate, which defines the problem of social justice.[10]

Now the original position is seen to be fair between equal moral persons once we grant that the natural distribution of abilities does not support a claim, grounded solely on an individual person's place in this distribution, to any particular scheme of background institutions, to a scheme that would favor that person's special endowments over the special endowments of others. This seems perfectly obvious. The veil of ignorance reflects this idea by excluding all knowledge of these matters in the original position. Neither the more nor the less fortunate as such has a claim to be especially favored. The basic structure, and the entitlements it generates and the legitimate expectations it honors, are to be governed by principles of justice that the parties adopt as representatives of free and equal moral persons.

In justice as fairness, then, there is no prior and independent notion of desert, perfectionist or intuitionist, that could override or

[10] For a discussion of the basic structure as the first subject of justice, see "The Basic Structure as Subject," in A. I. Goldman and Jaegwon Kim, eds., *Values and Morals* (Boston: Reidel, 1978), pp. 47–72.

restrict the agreement of the parties as agents of construction. To suppose that there is such a notion would violate the equality and autonomy of free and equal moral persons, which the rational autonomy of the parties in part represents. Thus citizens come to deserve this or that by their actual decisions and efforts within an on-going background system of cooperation with publicly announced rules that support legitimate expectations and acquired claims.[11] The only available notion of desert for judging this background system is derivative from the principles agreed to by the parties. Once this is recognized, the original position is seen as fair—or more accurately, as fair within a Kantian view, given its conception of free and equal persons, and of their autonomy.

VI

As I did in the first lecture, I conclude with a few general remarks. First, the guiding idea in representing persons is that, so far as possible, the parties in the original position as agents of construction should be constrained or influenced in the adoption of principles solely by features that fall under the Reasonable and the Rational and reflect the freedom and equality of moral persons. The original position thereby serves to connect, in the most explicit possible manner, the way the members of a well-ordered society view themselves as citizens with the content of their public conception of justice.

Another observation is that, although I regard justice as fairness as a Kantian view, it differs from Kant's doctrine in important respects. Here I note that justice as fairness assigns a certain primacy to the social; that is, the first subject of justice is the basic structure of society, and citizens must arrive at a public understanding on a conception of justice for this subject first. This understanding is interpreted via the unanimous agreement of the parties in the original position. By contrast, Kant's account of the Categorical Imperative applies to the personal maxims of sincere and conscientious individuals in everyday life. To be sure, in the course of testing such maxims we are to compare social worlds, that is, the social world that results when everyone follows the proposed maxim, as if by a law of nature, with the social world in which the contradictory maxim is followed. But this comparison of social worlds is undertaken singly by each person and for the purpose of judging a given personal maxim. Thus Kant proceeds from the particular, even personal, case of everyday life; he assumed that this process carried out correctly would eventually yield a coherent and suffi-

[11] See TJ, §48, and pp. 88 f.

ciently complete system of principles, including principles of social justice. Justice as fairness moves in quite the reverse fashion: its construction starts from a unanimous collective agreement regulating the basic structure of society within which all personal and associational decisions are to be made in conformity with this prior undertaking.

Finally, I have stressed the full publicity condition and its consequences for a conception of justice. Now this condition is related to the wide, as opposed to the narrow, view of the social role of morality.[12] The narrow view restricts this role to achieving the more or less minimum conditions of effective social cooperation, for example, by specifying standards to settle competing claims and setting up rules for coordinating and stabilizing social arrangements. Moral norms are regarded as inhibiting self- or group-centered tendencies, and aimed at encouraging less limited sympathies. Any moral doctrine accepts these requirements in some form, but they do not involve the full publicity condition. Once this condition is imposed, a moral conception assumes a wide role as part of public culture. Not only are its first principles embodied in political and social institutions and public traditions of their interpretation, but the derivation of citizens' rights, liberties, and opportunities invokes a certain conception of their person. In this way citizens are made aware of and educated to this conception. They are presented with a way of regarding themselves that otherwise they would most likely never have been able to entertain. Thus the realization of the full publicity condition provides the social milieu within which the notion of full autonomy can be understood and within which its ideal of the person can elicit an effective desire to be that kind of person. This educative role of the moral conception defines the wide view.

Now Kant often notes the publicity requirement in some form, but he seems to think that the conception of ourselves as fully autonomous is already given to us by the Fact of Reason, that is, by our recognition that the moral law is supremely authoritative for us as reasonable and rational beings.[13] Thus this conception of ourselves is implicit in individual moral consciousness, and the background social conditions for its realization are not emphasized or

[12] These terms are suggested by a similar distinction drawn by J. L. Mackie, *Ethics* (New York: Penguin Books, 1977), pp. 106 f, 134 ff.

[13] It is in some such fashion that I am inclined to interpret the important although difficult passages in the *Second Critique* where the Fact of Reason enters in.

made part of the moral doctrine itself. Justice as fairness departs from Kant, then, both in the primacy it assigns to the social and in the further aspect of this primacy contained in the full publicity condition. I believe these departures enable justice as fairness to avoid some of the faults that Dewey found in Kant's view.

CONSTRUCTION AND OBJECTIVITY

IN the preceding lectures I sketched the main idea of Kantian constructivism, which is to establish a connection between the first principles of justice and the conception of moral persons as free and equal. These first principles are used to settle the appropriate understanding of freedom and equality for a modern democratic society. The requisite connection is provided by a procedure of construction in which rationally autonomous agents subject to reasonable conditions agree to public principles of justice. With the sketch of these ideas behind us, I consider in this final lecture how a Kantian doctrine interprets the notion of objectivity in terms of a suitably constructed social point of view that is authoritative with respect to all individual and associational points of view. This rendering of objectivity implies that, rather than think of the principles of justice as true, it is better to say that they are the principles most reasonable for us, given our conception of persons as free and equal, and fully cooperating members of a democratic society. [Here 'reasonable' is used, as explained later (569/70), in contrast with 'true' as understood in rational intuitionism, and not, as previously (528–530), with 'rational', as in the notion of rational autonomy.]

I

To fix ideas, let's look back roughly a hundred years to Henry Sidgwick. *The Methods of Ethics* (first edition 1874) is, I believe, the outstanding achievement in modern moral theory.[1] By "moral theory" I mean the systematic and comparative study of moral conceptions, starting with those which historically and by current estimation seem to be the most important. Moral philosophy includes moral theory, but takes as its main question justification and how it is to be conceived and resolved; for example, whether it is to be conceived as an epistemological problem (as in rational intuitionism) or as a practical problem (as in Kantian constructivism). Sidgwick's *Methods* is the first truly academic work in moral theory,

[1] On Sidgwick now, see the comprehensive work by J. B. Schneewind, *Sidgwick's Ethics and Modern Victorian Moral Philosophy* (New York: Oxford, 1977).

modern in both method and spirit. Treating ethics as a discipline to be studied like any other branch of knowledge, it defines and carries out in exemplary fashion, if not for the first time, some of the comprehensive comparisons that constitute moral theory. By pulling together the work of previous writers, and through its influence on G. E. Moore and others, this work defined much of the framework of subsequent moral philosophy. Sidgwick's originality lies in his conception and mode of presentation of the subject and in his recognition of the significance of moral theory for moral philosophy.

It is natural, then, that the limitations of *Methods* have been as important as its merits. Of these limitations I wish to mention two. First, Sidgwick gives relatively little attention to the conception of the person and the social role of morality as main parts of a moral doctrine. He starts with the idea of a method of ethics as a method specified by certain first principles, principles by which we are to arrive at a judgment about what we ought to do. He takes for granted that these methods aim at reaching true judgments that hold for all rational minds. Of course, he thinks it is best to approach the problem of justification only when a broad understanding of moral theory has been achieved. In the preface of the first edition of *Methods* he explains that he wants to resist the natural urgency to discover the true method of ascertaining what it is right to do. He wishes instead to expound, from a neutral position and as impartially as possible, the different methods found in the moral consciousness of humankind and worked into familiar historical systems.[2] But these detailed expositions—necessary as they are—are merely preparation for comparing the various methods and evaluating them by criteria that any rational method that aims at truth must satisfy.

But a consequence of starting with methods of ethics defined as methods that seek truth is not only that it interprets justification as an epistemological problem, but also that it is likely to restrict attention to the first principles of moral conceptions and how they can be known. First principles are however only one element of a moral conception; of equal importance are its conception of the person and its view of the social role of morality. Until these other elements are clearly recognized, the ingredients of a constructivist doctrine are not at hand. It is characteristic of Sidgwick's *Methods* that the social role of morality and the conception of the person receive

2 *The Methods of Ethics* (London: Macmillan 1907), 7th ed., pp. v–vi; parenthetical page references to Sidgwick are to this book, this edition.

little notice. And so the possibility of constructivism was closed
to him.

Sidgwick overlooked this possibility because of a second limita-
tion: he failed to recognize that Kant's doctrine (and perfectionism
also for that matter) is a distinctive method of ethics. He regarded
the categorical imperative as a purely formal principle, or what he
called "the principle of equity": whatever is right for one person is
right for all similar persons in relevantly similar circumstances.
This principle Sidgwick accepts, but, since it is plainly not a suffi-
cient basis for a moral view, Kant's doctrine could not be counted
a substantive method (209/10). This formal reading of Kant, to-
gether with the dismissal of perfectionism, led Sidgwick to reduce
the traditional moral conceptions essentially to three main methods:
rational egoism, (pluralistic) intuitionism, and classical utilitarian-
ism. Surely he was right to restrict himself to a few conceptions so
that each could be explored in considerable detail. Only in this way
can depth of understanding be achieved. But rational egoism, which
he accepted as a method of ethics, is really not a moral conception
at all, but rather a challenge to all such conceptions, although no
less interesting for that. Left with only (pluralistic) intuitionism and
classical utilitarianism as methods of ethics in the usual sense, it is
no surprise that utilitarianism seemed superior to Sidgwick, given
his desire for unity and system in a moral doctrine.

Since Kant's view is the leading historical example of a construc-
tivist doctrine, the result once again is that constructivism finds no
place in *Methods*. Nor is the situation altered if we include another
leading representative work, F. H. Bradley's *Ethical Studies* (first
edition 1876); following Hegel, Bradley likewise regarded Kant's
ethics as purely formal and lacking in content and, therefore, to be
assigned to an early stage of the dialectic as an inadequate view.[3]
The result of these formal interpretations of Kant is that construc-
tivism was not recognized as a moral conception to be studied and
assimilated into moral theory. Nor was this lack made good in the
first half of this century; for in this period, beginning with Moore's
Principia Ethica (1903), interest centered mainly on philosophical
analysis and its bearing on justification regarded as an epistemolog-
ical problem and on the question whether its conclusions support
or deny the notion of moral truth. During this time, however, util-
itarianism and intuitionism made important advances. A proper
understanding of Kantian constructivism, on a par with our grasp
of these views, is still to be achieved.

[3] See Essay IV: "Duty for Duty's Sake," 2nd ed. (New York: Oxford, 1927).

II

Let us now try to deepen our understanding of Kantian constructivism by contrasting it with what I shall call *rational intuitionism*. This doctrine has, of course, been expressed in various ways; but in one form or another it dominated moral philosophy from Plato and Aristotle onwards until it was challenged by Hobbes and Hume, and, I believe, in a very different way by Kant. To simplify matters, I take rational intuitionism to be the view exemplified in the English tradition by Clarke and Price, Sidgwick and Moore, and formulated in its minimum essentials by W. D. Ross.[4] With qualifications, it was accepted by Leibniz and Wolff in the guise of perfectionism, and Kant knows of it in this form.

For our purposes here, rational intuitionism may be summed up by two theses: first, the basic moral concepts of the right and the good, and the moral worth of persons, are not analyzable in terms of nonmoral concepts (although possibly analyzable in terms of one another); and, second, first principles of morals (whether one or many), when correctly stated, are self-evident propositions about what kinds of considerations are good grounds for applying one of the three basic moral concepts, that is, for asserting that something is (intrinsically) good, or that a certain action is the right thing to do, or that a certain trait of character has moral worth. These two theses imply that the agreement in judgment which is so essential for an effective public conception of justice is founded on the recognition of self-evident truths about good reasons. And what these reasons are is fixed by a moral order that is prior to and independent of our conception of the person and the social role of morality. This order is given by the nature of things and is known, not by sense, but by rational intuition. It is with this idea of moral truth that the idea of first principles as reasonable will be contrasted.

It should be observed that rational intuitionism is compatible with a variety of contents for the first principles of a moral conception. Even classical utilitarianism, which Sidgwick was strongly inclined to favor (although he could not see how to eliminate rational egoism as a rival) was sometimes viewed by him as following from

4 See *The Right and the Good* (Oxford: The Clarendon Press, 1930), esp. chs. 1–2. I shall adopt Ross's characterization of rational intuitionism, adjusted to allow for any number of first principles and, thus, as fitting either single-principle or pluralistic intuitionism. I should add that, for my purposes here, I interpret Aristotle's view as combining teleological and metaphysical perfectionism. Although this may not be a sound interpretation in the light of contemporary scholarship, it suits well enough how Aristotle was interpreted up to Kant's time.

three principles each self-evident in its own right.[5] In brief, these three propositions were: the principle of equity so-called: that it cannot be right to treat two different persons differently merely on the ground of their being numerically different individuals; a principle of rational prudence: that mere difference of position in time is not by itself a reasonable ground for giving more regard to well-being at one moment than to well-being at another; and a principle of rational benevolence: the good of one person is of no more importance from the point of view of the universe than the good of any other person. These three principles, when combined with the principle that, as reasonable beings, we are bound to aim at good generally and not at any particular part of it, Sidgwick thought yielded the principle of utility: namely, to maximize the net balance of happiness. And this principle, like those from which it followed, he was tempted to hold as self-evident.

Of all recent versions of rational intuitionism, the appeal to self-evidence is perhaps most striking in Moore's so-called "ideal utilitarianism" in *Principia Ethica* (1903). A consequence of Moore's principle of organic unity is that his view is extremely pluralistic; there are few if any useful first principles, and distinct kinds of cases are to be decided by intuition as they arise. Moore held a kind of Platonic atomism: [6] moral concepts (along with other concepts) are subsisting and independent entities grasped by the mind. That pleasure and beauty are good, and that different combinations of them alone or together with other good things are also good, and to what degree, are truths known by intuition: by seeing with the mind's eye how these separate and distinct objects (universals) are (timelessly) related. This picture is even more vivid in the early philosophy of mathematics of Bertrand Russell, who talks of searching for the indefinable concepts of mathematics with a mental telescope (as one might look for a planet).[7]

Now my aim in recalling these matters is to point out that rational intuitionism, as illustrated by Sidgwick, Moore, and Ross, is sharply opposed to a constructivist conception along Kantian lines. That Kant would have rejected Hume's psychological naturalism as heteronomous is clear.[8] I believe that the contrast with

[5] *Methods*, Book III, ch. 13, pp. 379–389. See Schneewind's discussion, ch. 10, pp. 286–309.

[6] I borrow this expression from Peter Hylton's discussion, *The Origins of Analytic Philosophy*, ch. 3 (Dissertation: Harvard University, 1978).

[7] See *The Principles of Mathematics* (London: Allen & Unwin, 1937), 2nd ed. (1st ed. 1903), pp. xv–xvi. The analogy of the mental telescope is Russell's.

[8] Because it formulates definitions of the basic moral concepts in terms of

rational intuitionism, no matter what the content of the view (whether utilitarian, perfectionist, or pluralist) is even more instructive. It is less obvious that for Kant rational intuitionism is also heteronomous. The reason is that from the first thesis of rational intuitionism, the basic moral concepts are conceptually independent of natural concepts, and first principles are independent of the natural world and, as grasped by rational intuition, are regarded as synthetic a priori. This may seem to make these principles not heteronomous. Yet it suffices for heteronomy that these principles obtain in virtue of relations among objects the nature of which is not affected or determined by the conception of the person. Kant's idea of autonomy requires that there exist no such order of given objects determining the first principles of right and justice among free and equal moral persons. Heteronomy obtains not only when first principles are fixed by the special psychological constitution of human nature, as in Hume, but also when they are fixed by an order of universals or concepts grasped by rational intuition, as in Plato's realm of forms or in Leibniz's hierarchy of perfections.[9] Perhaps I should add, to prevent misunderstanding, that a Kantian doctrine of autonomy need not deny that the procedures by which first principles are selected are synthetic a priori. This thesis, however, must be properly interpreted. The essential idea is that such procedures must be suitably founded on practical reason, or, more exactly, on notions which characterize persons as reasonable and rational and which are incorporated into the way in which, as such

nonmoral concepts, this being the mode of identifying those facts which are to count as good reasons in applying the basic moral concepts, naturalism is a form of heteronomy from the Kantian standpoint. The various definitions, presumably arrived at by the analysis of concepts, convert moral judgments into statements about the world on all fours with those of science and common sense. Therefore, these definitions, combined with the natural order itself, now come to constitute the moral order, which is prior to and independent from our conception of ourselves as free and equal moral persons. If time permitted, this could be substantiated by setting out, for example, the details of Hume's view (as often interpreted) and of Bentham's hedonistic utilitarianism, at least once these views are expressed in the requisite naturalistic format. (Rational intuitionism tries to secure a kind of independence of the moral order from the order of nature.)

[9] This fundamental contention is unfortunately obscured by the fact that although in the *Grundlegung* Kant classifies the view of Leibniz and Wolff as a form of heteronomy, his criticism of it is that it is circular and therefore empty. See Academy Edition, p. 443. Much the same happens in the *Second Critique*, Academy Edition, p. 41, where Kant argues that the notion of perfection in practical reasoning means fitness for any given ends and therefore is again empty until these ends are specified independently. These arguments give the erroneous impression that, if perfectionism had sufficient content, it would be compatible with autonomy.

persons, they represent to themselves their free and equal moral personality. Put another way, first principles of justice must issue from a conception of the person through a suitable representation of that conception as illustrated by the procedure of construction in justice as fairness.

Thus in a Kantian doctrine a relatively complex conception of the person plays a central role. By contrast, rational intuitionism requires but a sparse notion of the person, founded on the self as knower. This is because the content of first principles is already fixed, and the only requirements on the self are to be able to know what these principles are and to be moved by this knowledge. A basic assumption is that the recognition of first principles as true and self-evident gives rise, in a being capable of rationally intuiting these principles, to a desire to act from them for their own sake. Moral motivation is defined by reference to desires that have a special kind of cause, namely, the intuitive grasp of first principles.[10] This sparse conception of the person joined with its moral psychology characterizes the rational intuitionism of Sidgwick, Moore, and Ross, although there is nothing that forces rational intuitionism to so thin a notion. The point is rather that, in rational intuitionism in contrast to a Kantian view, since the content of first principles is already given, a more complex conception of the person, of a kind adequate to determine the content of these principles, together with a suitable moral psychology, is simply unnecessary.

III

Having contrasted Kantian constructivism to rational intuitionism with respect to the idea of a moral order that is prior to and independent from our conception of the person, I now consider a second contrast, namely, how each regards the inevitable limitations that constrain our moral deliberations. The constructionist view accepts from the start that a moral conception can establish but a loose framework for deliberation which must rely very considerably on our powers of reflection and judgment. These powers are not fixed once and for all, but are developed by a shared public culture and hence shaped by that culture. In justice as fairness this means that the principles adopted by the parties in the original position are designed by them to achieve a public and workable agreement on matters of social justice which suffices for effective and fair social cooperation. From the standpoint of the parties as agents of construction, the first principles of justice are not thought to represent,

---

10 See, for example, *Methods*, pp. 23–28, 34–37, 39 f, read together with the discussion of the self-evident basis of the principle of utility, cited in fn 5 above.

or to be true of, an already given moral order, as rational intui-
tionism supposes. The essential point is that a conception of justice
fulfills its social role provided that citizens equally conscientious
and sharing roughly the same beliefs find that, by affirming the
framework of deliberation set up by it, they are normally led to a
sufficient convergence of opinion. Thus a conception of justice is
framed to meet the practical requirements of social life and to yield
a public basis in the light of which citizens can justify to one an-
other their common institutions. Such a conception need be only
precise enough to achieve this result.

On the constructivist view, the limitations that constrain our
moral deliberations affect the requirements of publicity and sup-
port the use of priority rules. These limitations also lead us to take
the basic structure of a well-ordered society as the first subject of
justice and to adopt the primary goods as the basis of interpersonal
comparisons. To begin with publicity: at the end of the preceding
lecture I mentioned why in a constructivist view first principles are
to satisfy the requirements of publicity. The moral conception is to
have a wide social role as a part of public culture and is to enable
citizens to appreciate and accept the conception of the person as
free and equal. Now if it is to play this wide role, a conception's
first principles cannot be so complex that they cannot be generally
understood and followed in the more important cases. Thus, it is
desirable that knowing whether these principles are satisfied, at
least with reference to fundamental liberties and basic institutions,
should not depend on information difficult to obtain or hard to
evaluate. To incorporate these desiderata in a constructivist view,
the parties are assumed to take these considerations into account
and to prefer (other things equal) principles that are easy to under-
stand and simple to apply. The gain in compliance and willing
acceptance by citizens more than makes up for the rough and ready
nature of the guiding framework that results and its neglect of cer-
tain distinctions and differences. In effect, the parties agree to rule
out certain facts as irrelevant in questions of justice concerning the
basic structure, even though they recognize that in regard to other
cases it may be appropriate to appeal to them. From the standpoint
of the original position, eliminating these facts as reasons of social
justice sufficiently increases the capacity of the conception to fulfill
its social role. Of course, we should keep in mind that the exclusion
of such facts as reasons of social justice does not alone entail that
they are not reasons in other kinds of situation where different
moral notions apply. Indeed, it is not even ruled out that the ac-

count of some notions should be constructivist, whereas the account of others is not.

It is evident, then, why a constructivist view such as justice as fairness incorporates into the framework of moral deliberation a number of schematic and practical distinctions as ways that enable us to deal with the inevitable limitations of our moral capacities and the complexity of our social circumstances. The need for such distinctions supports and helps to account for the use of certain priority rules to settle the relative weight of particular kinds of grounds in extremely important cases. Two such rules in justice as fairness are: first, the priority of justice over efficiency (in the sense of Pareto) and the net balance of advantages (summed over all individuals in society), and second, the priority of the principle of equal liberty (understood in terms of certain enumerated basic liberties) over the second principle of justice.[11] These rules are introduced to handle the complexity of the many prima facie reasons we are ready to cite in everyday life; and their plausibility depends in large part on the first principles to which they are adjoined. But although these rules are intended to narrow the scope of judgment in certain fundamental questions of justice, this scope can never be entirely eliminated, and for many other questions sharp and definite conclusions cannot usually be derived. Sharp and definite conclusions are not needed, however, if sufficient agreement is still forthcoming (TJ 44/5).

Similar considerations apply in beginning with the basic structure of a well-ordered society as the first subject of justice and trying to develop a conception of justice for this case alone. The idea is that this structure plays a very special role in society by establishing what we may call *background justice;* and if we can find suitable first principles of background justice, we may be able to exclude enough other considerations as irrelevant for this case, so as to develop a reasonably simple and workable conception of justice for the basic structure. The further complexities of everyday cases that cannot be ignored in a more complete moral conception may be dealt with later in the less general situations that occur within the various associations regulated by the basic structure, and in that sense subordinate to it.[12]

Finally, parallel observations hold in finding a feasible basis for

11 For a statement of these principles and priority rules, see TJ, pp. 60–62, 250, 302/3.

12 See "The Basic Structure as Subject," in A. I. Goldman and Jaegwon Kim, eds., *Values and Morals* (Boston: Reidel, 1978), especially secs. IV–V, pp. 52–57.

interpersonal comparisons of well-being relevant for questions of justice that arise in regard to the basic structure. These comparisons are to be made in terms of primary goods (as defined in the first lecture), which are, so far as possible, certain public features of social institutions and of people's situations with respect to them, such as their rights, liberties, and opportunities, and their income and wealth, broadly understood. This has the consequence that the comparison of citizens' shares in the benefits of social cooperation is greatly simplified and put on a footing less open to dispute.

Thus the reason why a constructivist view uses the schematic or practical distinctions we have just noted is that such distinctions are necessary if a workable conception of justice is to be achieved. These distinctions are incorporated into justice as fairness through the description of the parties as agents of construction and the account of how they are to deliberate. Charged with the task of agreeing to a workable conception of justice designed to achieve a sufficient convergence of opinion, the parties can find no better way in which to carry out this task. They accept the limitations of human life and recognize that at best a conception of justice can establish but a guiding framework for deliberation.

A comparison with classical utilitarianism will highlight what is involved here. On that view, whether stated as a form of rational intuitionism (Sidgwick) or as a form of naturalism (Bentham), every question of right and justice has an answer: whether an institution or action is right depends upon whether it will produce the greatest net balance of satisfaction. We may never be in a position to know the answer, or even to come very near to it, but, granting that a suitable measure of satisfaction exists, there is an answer: a fact of the matter. Of course, utilitarianism recognizes the needs of practice: working precepts and secondary rules are necessary to guide deliberation and coordinate our actions. These norms may be thought of as devised to bring our actions as close as possible to those which would maximize utility, so far as this is feasible. But of course, such rules and precepts are not first principles; they are at best directives that when followed make the results of our conduct approximate to what the principle of utility enjoins. In this sense, our working norms are approximations to something given.

By contrast, justice as fairness, as a constructivist view, holds that not all the moral questions we are prompted to ask in everyday life have answers. Indeed, perhaps only a few of them can be settled by any moral conception that we can understand and apply. Practical limitations impose a more modest aim upon a reasonable concep-

tion of justice, namely, to identify the most fundamental questions of justice that can be dealt with, in the hope that, once this is done and just basic institutions established, the remaining conflicts of opinion will not be so deep or widespread that they cannot be compromised. To accept the basic structure as the first subject of justice together with the account of primary goods is a step toward achieving this more modest goal. But in addition, the idea of approximating to moral truth has no place in a constructivist doctrine: the parties in the original position do not recognize any principles of justice as true or correct and so as antecedently given; their aim is simply to select the conception most rational for them, given their circumstances. This conception is not regarded as a workable approximation to the moral facts: there are no such moral facts to which the principles adopted could approximate.

As we have just seen, the differences between constructivism and classical utilitarianism are especially sharp in view of the content of the principle of utility: it always yields an answer that we can at least verbally describe. With the rational (pluralistic) intuitionism of Ross, however, the contrast is less obvious, since Ross's list of self-evident prima facie principles that identify good reasons also specifies but a loose guiding framework of moral deliberation which shares a number of the features of the framework provided by constructivism. But though these resemblances are real, the underlying idea of Ross's view is still essentially different from constructivism. His pluralistic intuitionism rejects utilitarianism (even an ideal utilitarianism) as oversimplifying the given moral facts, especially those concerning the correct weight of special duties and obligations. The complexity of the moral facts in particular kinds of cases is said to force us to recognize that no family of first principles that we can formulate characterizes these facts sufficiently accurately to lead to a definite conclusion. Decision and judgment are almost always to some degree uncertain and must rest with "perception," [13] that is, with our intuitive estimate of where the greatest balance of prima facie reasons lies in each kind of case. And this perception is that of a balance of reasons each of which is given by an independent moral order known by intuition. The essential contrast with constructivism remains.

IV

Having examined several contrasts between Kantian constructivism and rational intuitionism, we are now in a position to take up a

___

[13] See *The Right and the Good*, pp. 41/2. Ross refers to Aristotle's remark: "The decision rests with perception" (*Nicomachean Ethics* 1109 b 23, 1126 b 4).

fundamental point suggested by the discussion so far: an essential feature of a constructivist view, as illustrated by justice as fairness, is that its first principles single out what facts citizens in a well-ordered society are to count as reasons of justice. Apart from the procedure of constructing these principles, there are no reasons of justice. Put in another way, whether certain facts are to count as reasons of justice and what their relative force is to be can be ascertained only on the basis of the principles that result from the construction. This connects with the use of pure procedural justice at the highest level. It is, therefore, up to the parties in the original position to decide how simple or complex the moral facts are to be, that is, to decide on the number and complexity of the principles that identify which facts are to be accepted as reasons of justice by citizens in society (see TJ 45). There is nothing parallel to this in rational intuitionism.

This essential feature of constructivism may be obscured by the fact that in justice as fairness the first principles of justice depend upon those general beliefs about human nature and how society works which are allowed to the parties in the original position. First principles are not, in a constructivist view, independent of such beliefs, nor, as some forms of rational intuitionism hold, true in all possible worlds. In particular, they depend on the rather specific features and limitations of human life that give rise to the circumstances of justice.[14] Now, given the way the original position is set up, we can allow, in theory, that, as the relevant general beliefs change, the beliefs we attribute to the parties likewise change, and conceivably also the first principles that would be agreed to. We can say, if we like that *the* (most reasonable) principles of justice are those which would be adopted if the parties possessed all relevant general information and if they were properly to take account of all the practical desiderata required for a workable public conception of justice. Though these principles have a certain pre-eminence, they are still the outcome of construction. Furthermore, it is important to notice here that no assumptions have been made about a theory of truth. A constructivist view does not require an idealist or a verificationist, as opposed to a realist, account of truth. Whatever the nature of truth in the case of general beliefs about human nature and how society works, a constructivist moral doctrine requires a distinct procedure of construction to identify the first principles of justice. To the extent that Kant's moral doctrine

14 See Lecture II, section I.

depends upon what to some may appear to be a constructivist account of truth in the *First Critique* (I don't mean to imply that such an interpretation is correct), justice as fairness departs from that aspect of Kant's view and seeks to preserve the over-all structure of his moral conception apart from that background.

In the preceding paragraph I said that the way justice as fairness is set up allows the possibility that, as the general beliefs ascribed to the parties in the original position change, the first principles of justice may also change. But I regard this as a mere possibility noted in order to explain the nature of a constructivist view. To elaborate: at the end of the first lecture I distinguished between the roles of a conception of the person and of a theory of human nature, and I remarked that in justice as fairness these are distinct elements and enter at different places. I said that a conception of the person is a companion moral ideal paired with the ideal of a well-ordered society. A theory of human nature and a view of the requirements of social life tell us whether these ideals are feasible, whether it is possible to realize them under normally favorable conditions of human life. Changes in the theory of human nature or in social theory generally which do not affect the feasibility of the ideals of the person and of a well-ordered society do not affect the agreement of the parties in the original position. It is hard to imagine realistically any new knowledge that should convince us that these ideals are not feasible, given what we know about the general nature of the world, as opposed to our particular social and historical circumstances. In fact, the relevant information on these matters must go back a long time and is available to the common sense of any thoughtful and reflective person. Thus such advances in our knowledge of human nature and society as may take place do not affect our moral conception, but rather may be used to implement the application of its first principles of justice and suggest to us institutions and policies better designed to realize them in practice.[15]

In justice as fairness, then, the main ideals of the conception of justice are embedded in the two model-conceptions of the person and of a well-ordered society. And, granting that these ideals are allowed by the theory of human nature and so in that sense feasible, the first principles of justice to which they lead, via the constructivist

[15] Therefore these advances in our knowledge of human psychology and social theory might be relevant at the constitutional, legislative, and judicial stages in the application of the principles of justice, as opposed to the adoption of principles in the original position. For a brief account of these stages, see TJ, §31.

procedure of the original position, determine the long-term aim of social change. These principles are not, as in rational intuitionism, given by a moral order prior to and independent from our conception of the person and the social role of morality; nor are they, as in some naturalist doctrines, to be derived from the truths of science and adjusted in accordance with advances in human psychology and social theory. (These remarks are admittedly too brief, but we must return to the main line of discussion.)

V

The rational intuitionist may object that an essential feature of constructivism—the view that the facts to count as reasons of justice are singled out by the parties in the original position as agents of construction and that, apart from such construction, there are no reasons of justice—is simply incoherent.[16] This view is incompatible not only with the notion of truth as given by a prior and independent moral order, but also with the notions of reasonableness and objectivity, neither of which refer to what can be settled simply by agreement, much less by choice. A constructivist view, the objection continues, depends on the idea of adopting or choosing first principles, and such principles are not the kind of thing concerning which it makes sense to say that their status depends on their being chosen or adopted. We cannot "choose" them; what we can do is choose whether to follow them in our actions or to be guided by them in our reasoning, just as we can choose whether to honor our duties, but not what our duties are.

In reply, one must distinguish the three points of view that we noted at the end of the first lecture (in section VII, 533/4): that of the parties in the original position, that of the citizens in a well-ordered society, and that of you and me who are examining justice as fairness to serve as a basis for a conception that may yield a suitable understanding of freedom and equality. It is, of course, the parties in the original position whose agreement singles out the facts to count as reasons. But their agreement is subject to all the conditions of the original position which represent the Reasonable and the Rational. And the facts singled out by the first principles count as reasons not for the parties, since they are moved by their highest-

16 For this and other objections to what I call "constructivism" in this lecture, see the review of TJ by Marcus Singer, *Philosophy of Science*, XLIV, 4 (December 1977): 594–618, pp. 612–615. I am grateful to him for raising this objection, which I here try to meet. Singer's criticism starts from the passage on page 45 of TJ (also referred to above, 564/5). It should not be assumed that Singer's own position is that of rational intuitionism. I simply suppose that a rational intuitionist would make this objection.

order interests, but for the citizens of a well-ordered society in matters of social justice. As citizens in society we are indeed bound by first principles and by what our duties are, and must act in the light of reasons of justice. Constructivism is certain to seem incoherent unless we carefully distinguish these points of view.

The parties in the original position do not agree on what the moral facts are, as if there already were such facts. It is not that, being situated impartially, they have a clear and undistorted view of a prior and independent moral order. Rather (for constructivism), there is no such order, and therefore no such facts apart from the procedure of construction as a whole; the facts are identified by the principles that result. Thus the rational intuitionists' objection, properly expressed, must be that no hypothetical agreement by rationally autonomous agents, no matter how circumscribed by reasonable conditions in a procedure of construction, can determine the reasons that settle what we as citizens should consider just and unjust; right and wrong are not, even in that way, constructed. But this is merely to deny what constructivism asserts. If, on the other hand, such a construction does yield the first principles of a conception of justice that matches more accurately than other views our considered convictions in general and wide reflective equilibrium, then constructivism would seem to provide a suitable basis for objectivity.

The agreement of the parties in the original position is not a so-called "radical" choice: that is, a choice not based on reasons, a choice that simply fixes, by sheer fiat, as it were, the scheme of reasons that we, as citizens, are to recognize, at least until another such choice is made. The notion of radical choice, commonly associated with Nietzsche and the existentialists, finds no place in justice as fairness. The parties in the original position are moved by their preference for primary goods, which preference in turn is rooted in their highest-order interests in developing and exercising their moral powers. Moreover, the agreement of the parties takes place subject to constraints that express reasonable conditions.

In the model-conception of a well-ordered society, citizens affirm their public conception of justice because it matches their considered convictions and coheres with the kind of persons they, on due reflection, want to be. Again, this affirmation is not radical choice. The ideals of the person and of social cooperation embedded in the two model-conceptions mediated by the original position are not ideals that, at some moment in life, citizens are said simply to choose. One is to imagine that, for the most part, they find on exam-

ination that they hold these ideals, that they have taken them in part from the culture of their society.

The preceding paragraph ties in with what I said at the beginning of the first lecture, except that there I was talking about us and not about a well-ordered society. Recall that a Kantian view, in addressing the public culture of a democratic society, hopes to bring to awareness a conception of the person and of social cooperation conjectured to be implicit in that culture, or at least congenial to its deepest tendencies when properly expressed and presented. Our society is not well-ordered: the public conception of justice and its understanding of freedom and equality are in dispute. Therefore, for us—you and me—a basis of public justification is still to be achieved. In considering the conception of justice as fairness we have to ask whether the ideals embedded in its model-conceptions are sufficiently congenial to our considered convictions to be affirmed as a practicable basis of public justification. Such an affirmation would not be radical choice (if choice at all); nor should it be confused with the adoption of principles of justice by the parties in the original position. To the contrary, it would be rooted in the fact that this Kantian doctrine as a whole, more fully than other views available to us, organized our considered convictions.

Given the various contrasts between Kantian constructivism and rational intuitionism, it seems better to say that in constructivism first principles are reasonable (or unreasonable) than that they are true (or false)—better still, that they are most reasonable for those who conceive of their person as it is represented in the procedure of construction. And here 'reasonable' is used instead of 'true' not because of some alternative theory of truth, but simply in order to keep to terms that indicate the constructivist standpoint as opposed to rational intuitionism. This usage, however, does not imply that there are no natural uses for the notion of truth in moral reasoning. To the contrary, for example, particular judgments and secondary norms may be considered true when they follow from, or are sound applications of, reasonable first principles. These first principles may be said to be true in the sense that they would be agreed to if the parties in the original position were provided with all the relevant true general beliefs.

Nor does justice as fairness exclude the possibility of there being a fact of the matter as to whether there is a single most reasonable conception. For it seems quite likely that there are only a few viable conceptions of the person both sufficiently general to be part of a moral doctrine and congruent with the ways in which people are

to regard themselves in a democratic society. And only one of these conceptions may have a representation in a procedure of construction that issues in acceptable and workable principles, given the relevant general beliefs.[17] Of course, this is conjecture, intended only to indicate that constructivism is compatible with there being, in fact, only one most reasonable conception of justice, and therefore that constructivism is compatible with objectivism in this sense. However, constructivism does not presuppose that this is the case, and it may turn out that, for us, there exists no reasonable and workable conception of justice at all. This would mean that the practical task of political philosophy is doomed to failure.

VI

My account of Kantian constructivism in moral theory (as illustrated by justice as fairness) is now concluded. I should stress, however, that for all I have said it is still open to the rational intuitionist to reply that I have not shown that rational intuitionism is false or that it is not a possible basis for the necessary agreement in our judgments of justice. It has been my intention to describe constructivism by contrast and not to defend it, much less to argue that rational intuitionism is mistaken. In any case, Kantian constructivism, as I would state it, aims to establish only that the rational intuitionist notion of objectivity is unnecessary for objectivity. Of course, it is always possible to say, if we ever do reach general and wide reflective equilibrium, that now at last we intuit the moral truths fixed by a given moral order; but the constructivist will say instead that our conception of justice, by all the criteria we can think of to apply, is now the most reasonable for us.

We have arrived at the idea that objectivity is not given by "the point of view of the universe," to use Sidgwick's phrase. Objectivity is to be understood by reference to a suitably constructed social point of view, an example of which is the framework provided by the procedure of the original position. This point of view is social in several respects. It is the publicly shared point of view of citizens in a well-ordered society, and the principles that issue from it are accepted by them as authoritative with regard to the claims of individuals and associations. Moreover, these principles regulate the basic structure of society within which the activities of individuals and associations take place. Finally, by representing the person as a free and equal citizen of a well-ordered society, the constructivist procedure yields principles that further everyone's highest-order interests and define the fair terms of social cooperation among persons

17 I am indebted to Samuel Scheffler for valuable discussion on this point.

so understood. When citizens invoke these principles they speak as members of a political community and appeal to its shared point of view either in their own behalf or in that of others. Thus, the essential agreement in judgments of justice arises not from the recognition of a prior and independent moral order, but from everyone's affirmation of the same authoritative social perspective.

The central place of the conception of the person in these lectures prompts me to conclude with a note of warning, addressed as much to me as to anyone else: ever since the notion of the person assumed a central place in moral philosophy in the latter part of the eighteenth century, as seen in Rousseau and Kant and the philosophy of idealism, its use has suffered from excessive vagueness and ambiguity. And so it is essential to devise an approach that disciplines our thought and suitably limits these defects. I view the three model-conceptions that underlie justice as fairness as having this purpose.

To elucidate: suppose we define the concept of a person as that of a human being capable of taking full part in social cooperation, honoring its ties and relationships over a complete life. There are plainly many specifications of this capacity, depending, for example, on how social cooperation or a complete life is understood; and each such specification yields another conception of the person falling under the concept. Moreover, such conceptions must be distinguished from specifications of the concept of the self as knower, used in epistemology and metaphysics, or the concept of the self as the continuant carrier of psychological states: the self as substance, or soul. These are prima facie distinct notions, and questions of identity, say, may well be different for each; for these notions arise in connection with different problems. This much is perhaps obvious. The consequence is that there are numerous conceptions of the person as the basic unit of agency and responsibility in social life, and of its requisite intellectual, moral, and active powers. The specification of these conceptions by philosophical analysis alone, apart from any background theoretical structure or general requirements, is likely to prove fruitless. In isolation these notions play no role that fixes or limits their use, and so their features remain vague and indeterminate.

One purpose of a model-conception like that of the original position is that, by setting up a definite framework within which a binding agreement on principles must be made, it serves to fix ideas. We are faced with a specific problem that must be solved, and we are forced to describe the parties and their mutual relations

in the process of construction so that appropriate principles of justice result. The context of the problem guides us in removing vagueness and ambiguity in the conception of the person, and tells us how precise we need to be. There is no such thing as absolute clarity or exactness; we have to be only clear or exact enough for the task at hand. Thus the structure defined by the original position may enable us to crystallize our otherwise amorphous notion of the person and to identify with sufficient sharpness the appropriate characterization of free and equal moral personality.

The constructivist view also enables us to exploit the flexibility and power of the idea of rational choice subject to appropriate constraints. The rational deliberations of the parties in the original position serve as a way to select among traditional or other promising conceptions of justice. So understood, the original position is not an axiomatic (or deductive) basis from which principles are derived but a procedure for singling out principles most fitting to the conception of the person most likely to be held, at least implicitly, in a modern democratic society. To exaggerate, we compute via the deliberations of the parties and in this way hope to achieve sufficient rigor and clarity in moral theory. Indeed, it is hard to see how there could be any more direct connection between the conception of free and equal moral persons and first principles of justice than this construction allows. For here persons so conceived and moved by their highest-order interests are themselves, in their rationally autonomous deliberations, the agents who select the principles that are to govern the basic structure of their social life. What connection could be more intimate than this?

Finally, if we ask, what is clarity and exactness enough? the answer is: enough to find an understanding of freedom and equality that achieves workable public agreement on the weight of their respective claims. With this we return to the current impasse in the understanding of freedom and equality which troubles our democratic tradition and from which we started. Finding a way out of this impasse defines the immediate practical task of political philosophy. Having come full circle, I bring these lectures to a close.

JOHN RAWLS

Harvard University

NOTES AND NEWS

L'Association des Etudes grecques announces the award of its Théodore Reinach Prize to Janine Chanteur (Université de Paris—Sorbonne) for her