



## Rawls' Theory of Justice--I

R. M. Hare

*The Philosophical Quarterly*, Vol. 23, No. 91 (Apr., 1973), 144-155.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8094%28197304%2923%3A91%3C144%3ARTOJ%3E2.0.CO%3B2-W>

*The Philosophical Quarterly* is currently published by The Philosophical Quarterly.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/philquar.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## CRITICAL STUDY

RAWLS' THEORY OF JUSTICE—I<sup>1</sup>

BY R. M. HARE

*A Theory of Justice.* By JOHN RAWLS. (Cambridge, Mass. : Harvard U.P., 1971. Oxford : Clarendon Press, 1972. Pp. xv + 607. Price £5.00, U.S. paperback, \$3.95.)

Any philosopher who writes on justice or on any other subject in moral philosophy is likely to propound, or to give evidence of, views on one or more of the following topics :

- (1) *Philosophical methodology*—i.e., what philosophy is supposed to be doing and how it does it. Rawls expresses some views about this, which have determined the whole structure of his argument, and which therefore need careful inspection.
- (2) *Ethical analysis*—i.e., the meanings of the moral words or the nature and logical properties of the moral concepts. Rawls says very little about these, and certainly does not treat them as fundamental to his enquiry (51/10)<sup>2</sup>.
- (3) *Moral methodology*—i.e., how moral thinking ought to proceed, or how moral arguments or reasonings have to be conducted if they are to be cogent.
- (4) *Normative moral questions*—i.e., what we ought or ought not to do, what is just or unjust, and so on.

I shall leave discussion of Rawls' views on (4) to the second, forthcoming part of this review, this first part being devoted to (1), (2) and (3). I shall argue that, through misconceptions about (1), Rawls has not paid enough attention to (2), and that therefore he has lacked the equipment necessary to handle (3) effectively ; so that what he says about (4), however popular it may prove, is unsupported by any firm arguments.

(1) Rawls states quite explicitly how he thinks moral philosophy should be done : " There is a definite if limited class of facts against which conjec-

<sup>1</sup>Although the Editor has been kind enough to allow me to spread this review over two parts, I do not hope to explore in it all the convolutions of the book. I shall concentrate on what seems most important. I feel excused from discussing Rawls' treatment of liberty by my general agreement with an article which Professor Hart is to devote to this topic in the *Chicago Law Review*, and of which he has kindly shown me a draft. Of the many other people with whom I have discussed the book, and who have kept my courage up during two readings of it, I should like especially to thank Mr. Derek Parfit, who seems to me to see deeper and more clearly into these problems than any of us. [The second part of this Critical Study is to appear in the July number of this volume.]

<sup>2</sup>References are to pages/lines of Rawls' text.

tured principles can be checked, namely our considered judgments in reflective equilibrium" (51/3). It is clear from the succeeding passage that Rawls does not conceive of moral philosophy as depending primarily on the analysis of concepts in order to establish their logical properties and thus the rules of valid moral argument. Rather, he thinks of a theory of justice as analogous to a theory in empirical science. It has to square with what he calls "facts", just like, for example, physiological theories. But what are these facts? They are what people will *say* when they have been thinking carefully. This suggestion is reminiscent of Sir David Ross.<sup>3</sup> But sometimes (though not consistently) Rawls goes farther than Ross. Usually he is more cautious, and appeals to the reflections of *bien pensants* generally, as Ross does (e.g., 18/9, 19/26). But at 50/34 he says, "For the purposes of this book, the views of the reader and the author are the only ones that count". It does not make much practical difference which way he puts it; for if (as will certainly be the case) he finds a large number of readers who can share with him a cosy unanimity in their considered judgments, he and they will think that they adequately represent "people generally", and congratulate themselves on having attained the truth.<sup>4</sup> This is how phrases like 'reasonable and generally acceptable' (45/16) are often used by philosophers in lieu of argument.

Rawls, in short, is here advocating a kind of subjectivism, in the narrowest and most old-fashioned sense. He is making the answer to the question "Am I right in what I say about moral questions?" depend on the answer to the question "Do you, the reader, and I agree in what we say?". This must be his view, if the considered judgments of author and reader are to occupy the place in his theory which is occupied in an empirical science by the facts of observation. Yet at 516/15 he claims objectivity for his principles.

It might be thought that such a criticism can be made only by one who has rejected (as Rawls has apparently accepted) the arguments of Professor Quine and others about the analytic-synthetic distinction and the way in which science confronts the world. But this is not so. Even Quine would hardly say that scientific theories as a whole are to be tested by seeing what people say when they have thought about them (it would have been a good thing for medieval flat-earthers if they could be); but that is what Rawls is proposing for moral principles.

In order not to be unfair to Rawls, it must be granted that *any* enquirer, in ethics as in any other subject, and whether he be a descriptivist or a prescriptivist, is looking for an answer to his questions which he can accept. I have myself implied this in my *Freedom and Reason*, page 73 and elsewhere. The element of subjectivism enters only when a philosopher claims that he can "check" his theory against his and other people's views, so

<sup>3</sup>Cf. *The Right and the Good*, pp. 40 ff.

<sup>4</sup>See 104/3-14 for a "considered judgment" with which many of us now would agree, but which differs from the views of most writers of other periods than the present, and is not argued for.

that a disagreement between the theory and the views tells against the theory. To speak like this (as Rawls does constantly throughout the book) is to make the *truth* of the theory *depend on* agreement with people's opinions. I have myself been so often falsely accused of this sort of subjectivism that it is depressing to find a self-styled objectivist falling as deeply into it as Rawls does—depressing, because it makes one feel that this essentially simple distinction will never be understood: the distinction between the view that thinking something can make it so (which is in general false) and the view that if we are to say something sincerely, we must be able to accept it (which is a tautology).

Intuitionism is nearly always a form of disguised subjectivism. Rawls does not call himself an intuitionist; but he certainly is one in the usual sense. He says, "There is no reason to suppose that we can avoid all appeals to intuition, of whatever kind, or that we should try to. The practical aim is to reach a reasonably reliable agreement in judgment in order to provide a common conception of justice" (44/34, cf. 124/38). It is clear that he is here referring mainly to moral intuitions; perhaps if he appealed only to linguistic intuitions it would be all right. He reserves the name 'intuitionist' for those (including no doubt Ross) who advocate a *plurality* of moral principles, each established by intuition, and not related to one another in an ordered structure, but only weighed relatively to each other (also by intuition) when they conflict. The right name for this kind of intuitionism would be 'pluralistic intuitionism'. Rawls' theory is more systematic than this, but no more firmly grounded. There can also be another, non-pluralistic kind of intuitionist—one who intuits the validity of a single principle or ordered system of them, or of a single method, and erects his entire structure of moral thought on this. Sidgwick might come into this category—though if he were living today, it is unlikely that he would find it necessary to rely on moral intuition.

'Monistic intuitionism' would be a good description of this kind of view. It might apply to Rawls, did it not suggest falsely that he relies only on one great big intuition, and only at one point in his argument. Unfortunately he relies on scores of them. From 18/9 to 20/9 I have counted in two pages thirty expressions implying a reliance on intuitions: such expressions as 'I assume that there is a broad measure of agreement that'; 'commonly shared presumptions'; 'acceptable principles'; 'it seems reasonable to suppose'; 'is arrived at in a natural way'; 'match our considered convictions of justice or extend them in an acceptable way'; 'which we can affirm on reflection'; 'we are confident'; 'we think'; and so on. If, as I have done, the reader will underline the places in the book where crucial moves in the argument depend on such appeals, he may find himself recalling Plato's remark: "If a man starts from something he knows not, and the end and middle of his argument are tangled together out of what he knows not, how can such a mere consensus ever turn into

knowledge?" (*Rep.* 533 c). Since the theoretical structure is tailored at every point to fit Rawls' intuitions, it is hardly surprising that its normative consequences fit them too—if they did not, he would alter the theory (19/28 ff., cf. 141/23); and the fact that Rawls is a fairly typical man of his times and society, and will therefore have many adherents, does not make this a good way of doing philosophy.

Rawls' answer to this objection (581/9) is that *any* justification of principles must proceed from some consensus. It is true that any justification which consists of a "linear inference"<sup>5</sup> must so proceed; but Rawls' justification is not of this type. Why should it not *end* in consensus as a result of argument? There may have to be a prior consensus on matters of fact, including facts about the interests of the parties (though these themselves may conflict); and on matters of logic, established by analysis. But not on substantial moral questions, as Rawls seems to require. A review is not the place for an exposition of my own views of how moral argument can succeed in reaching normative conclusions with only facts, *singular* prescriptions and logic to go on; all that I wish to say here is that the matter will never be clarified unless these ingredients are kept meticulously distinct, and the logic carefully attended to (see further footnote 6).

(2) I shall mention only in passing Rawls' views about the meanings of the moral words or the natures, analyses and logical properties of the moral concepts. It would be wrong to take up space on something which Rawls evidently thinks of little importance for his argument. He wishes to "leave questions of meaning and definition aside and to get on with the task of developing a substantive theory of justice" (579/17). There is in fact a vast hole in his 600-page book which should be occupied by a thorough account of the meanings of these words, which is the only thing that can establish the logical rules that govern moral argument. If we do not have such an account, we shall never be able to distinguish between what we have to avoid saying if we are not to contradict ourselves or commit other *logical* errors, and what we have to avoid saying if we are to agree with Rawls and his coterie.<sup>6</sup> So far as he does say anything about the meanings of the moral words, it is mostly derivative from recent descriptivist views, my arguments against which it would be tedious to rehearse. I found this reliance surprising, in view of the fact that what he says about justice, at any rate, clearly commits him to some form of prescriptivism: the principles of justice determine how we *are to* behave, not how we are to *describe* certain kinds of behaviour (61/7, 145/12, 14, 33, 149/16, 351/15). My quarrels with Rawls' main theory do not depend at all on the fact that I am a prescriptivist.

<sup>5</sup>See my *Freedom and Reason*, pp. 87 f.

<sup>6</sup>See the paper "The Argument from Received Opinion" in my *Essays on Philosophical Method*, which might have been written with Rawls' book in mind, although in fact at that time I had not had the opportunity of reading it. For my latest shot at the project of giving such an account, see the paper "Wrongness and Harm" in my *Essays on the Moral Concepts*.

There are significant passages in which Rawls compares moral philosophy with mathematics (51/23) and linguistics (47/5, 49/8). The analogy with these sciences is vitiated by the fact that they do not yield substantial conclusions, as moral philosophy is supposed, on Rawls' view, to do, and in some sense clearly should. It is quite all right to test a linguistic theory (a grammar) against what people actually say when they are speaking carefully; people's *linguistic* "intuitions" are indeed, in the end, authoritative for what is correct in their language. The kind of interplay between theory and data that occurs in all sciences can occur here, and it is perfectly proper for the data to be the utterances of native speakers. But the only "moral" theories that can be checked against people's actual moral judgments are anthropological theories about what, in general, people *think* one ought to do, not moral principles about what one ought to do. That these latter can be so checked is not, indeed, what Rawls is suggesting in this passage; but do not the whole drift of his argument, and the passage quoted above (51/3), suggest it?

The case of mathematics is more controversial. Rawls seems to imply that if we had a "moral system" analogous to the systems of logic and mathematics, then we could use such a system to elucidate the meanings of moral judgments, instead of the other way about, as I have suggested. There is no objection, so far as I can see, to such a claim in mathematics and logic, provided that we realize that the concepts used in the formal systems may be different from (perhaps more useful for certain purposes than) our natural ones. Such a procedure is all right in logic and mathematics, since the construction of artificial models can often illuminate the logic and the meaning of our ordinary speech; but whichever way the illumination goes (why not both ways?) it can work only if the system in question is purely formal. If what Rawls calls "the substantive content of the moral conceptions" (52/7) is part of the system, then what will be revealed by it are not the meanings of moral judgments but the moral opinions of those who adhere to the system. And when he proposes (111/6) to *replace* our concept of right by the concept of being in accordance with the principles that would be acknowledged in the original position, he is in effect seeking to foist on us not a new meaning for a word, but a substantial set of moral views; for he thinks that he has tailored the original position so as to yield principles which fit his own considered judgments.

(3) Rawls' moral methodology takes the form of a picture or parable—and one which is even more difficult than most to interpret with any confidence. We are to imagine a set of people gathered together (hypothetically, not actually), to agree upon a set of "principles of justice" to govern their conduct.<sup>7</sup> The "principles of justice" are those principles to which these

<sup>7</sup>It is tempting to say "their *subsequent* conduct"; but the tenses in Rawls' account are one of its most baffling features. On the one hand, these "people in the original position" (POPs) are to make a "contract"; this, and terms like 'original position'

“people in the original position” (POPs) would agree for the conduct of all of them as “people in ordinary life” (POLs), if, when making the agreement, they were subject to certain conditions.

It is obviously these conditions which determine the substance of the theory (indeed they *are* its substance, the rest being mere dramatization, useful for expository purposes, but also potentially misleading). Rawls' theory belongs to a class of theories which we may call “hypothetical choice theories”—i.e., theories which say that the right answer to some question is the answer that a person or set of people *would* choose if subject to certain conditions. The best-known example of such a theory is the “ideal observer” theory of ethics, about which Rawls says something, and which we shall find instructive to compare with his own. The important thing to notice about all such theories is that *what* this hypothetical person would choose, if it is determinate at all (which many such theories fail to make it) has to be determined by the conditions to which he is subject. If the conditions, once made explicit, do not deductively determine the choice, then the choice remains indeterminate, except in so far as it is covertly conditioned by the prejudices or intuitions of the philosopher whose theory it is. Thus intuition can enter at two points (and in Rawls' case enters at both; cf. 121/7-15). It enters in the choice of the conditions to which the chooser is to be subject; and it enters to determine what he will choose in cases where the conditions, as made explicit, do not determine this (see Part II, forthcoming).

The more important of the conditions to which Rawls' POPs are subject are the following (§§22-5):

- (1) They know certain facts about the world and the society in which POLs live, but have others concealed from them by a “veil of ignorance”. It is obviously going to be crucial *which* facts they are allowed to know, and which they are not.
- (2) They are motivated in certain ways, especially in being selfish or mutually disinterested, and also in lacking envy and in being unwilling to use the principle of insufficient reason. They are also “rational” (i.e., take the most effective means to given ends (14/5)).
- (3) They are subject to “the formal constraints of the concept of

and ‘initial situation’ (20/18), seem to indicate that this conclave is temporally prior to the time at which these same people are to enter the world as we know it, become “people in ordinary life” (POLs) and carry out their contract. But on the other hand Rawls seems to speak commonly in the present tense (e.g., 520/27 ff.), as if they were somehow simultaneously POPs and POLs. Not surprisingly this, and other obscurities, make it often difficult, and sometimes (to me at any rate) impossible, to determine whether some particular remark is intended to refer to POPs or POLs. Who, for example, are “they” in 206/5? And in 127/25, is it the POLs who are being said to be mutually disinterested, as the passage seems to imply, and as is suggested by the reference to “circumstances of justice” on 128/5 (which seems usually, though not on 130/1-5, to mean circumstances of POLs, not POPs)? But if so, how are 129/14-18 or 148/2 ff. consistent? Again, do POLs lack envy, or only POPs? (see 151/22-24, 143-4, §§ 80-81.) A review as long as Rawls' book itself could be spent on such questions of interpretation; I was intending to set a few more of these exercises for the reader, but have not had time to compile them.

right". Rawls explicitly says that he does not "claim that these conditions follow from the concept of right, much less from the meaning of morality" (130/16). Instead, he as usual says that it "seems reasonable" to impose them (130/14). He does not tell us what he would say to somebody to whom they did not "seem reasonable".

- (4) There are also certain important procedural stipulations, such as that the POPs should all agree unanimously in their choice of principles. Later in the book, the procedure is very much elaborated, and takes the form of a series of stages in which the "veil of ignorance" is progressively lifted; but I shall ignore this complication here.

In comparing Rawls' theory with other theories, it is most important to notice the roles played by these groups of conditions. If I may be allowed to mention my own theory, I would myself place almost the whole emphasis on (3), and would at the same time aim to establish the "constraints" on the basis of a study of the logical properties of the moral words. This still seems to me the most rigorous and secure procedure, because it enables us to say that *if* this is how we are using the words (*if* this is what we mean by them), then we shall be debarred from saying so-and-so on pain of self-contradiction; and this gives moral arguments a cutting edge which in Rawls they lack. In a similar way, Achilles should have answered the Tortoise by saying, "If you mean by 'if' what we all mean, you have to accept *modus ponens*; for this is the rule that gives its meaning to 'if'". It is of course in dispute how much we can do by this method; but I think, and have tried elsewhere to show, that we can do much more than Rawls allows.

The "ideal observer" theory (in a typical form) differs from Rawls' theory in the following respects. Under (1), it allows the principle-chooser to know everything; there is no "veil of ignorance". On the other hand, under (2), he is differently motivated; instead of being concerned with his own interest only, he is impartially benevolent. Now it is possible to show that on a certain simple and natural "rational contractor" theory of the Rawls type (though not, it is fairly safe to say, on Rawls' own version of this type of theory) these two changes exactly cancel one another, so that the normative consequences of the "ideal observer" and "rational contractor" theories would be identical. To see this, let us remember that the main object of these conditions is to secure impartiality. This is secured in the case of the rational contractor theory by not allowing the POPs to know what are to be their individual roles as POLs in the society in which the contract has to be observed; they therefore cannot choose the principles to suit their own selfish interests, although they are selfishly motivated. It is secured in the case of the ideal observer theory by express stipulation;



he is required to be impartially benevolent. It looks, therefore, as if *these* versions of the two theories are, as I have said elsewhere,<sup>9</sup> practically equivalent.

We must next ask *how much* the POPs have to be ignorant of, in order to secure impartiality. It must be noticed that much of the work is already done by the "formal constraint" that the principles have to be "general".<sup>9</sup> Rawls himself says that the formal constraints rule out egoism (136/13); it might therefore be asked what there is left for the "veil of ignorance" to do, since to abandon egoism (and for the same formal reasons the pursuit of the interests of any other particular person or set of them) is *eo ipso* to become impartial. I do not think that this objection sticks; for a POP, if he had full knowledge of his own role as a POL, might adopt principles which were formally "general" or universal but were rigged to suit his own interest. Rawls, however, thinks (wrongly<sup>10</sup>) that such rigged principles can be ruled out on the formal ground of lack of "generality", and so is open to the objection *ad hominem*. That is to say, *he* has left nothing for the veil of ignorance to do as regards impartiality.

Be that as it may, however, we need to be clear how thick a veil of ignorance is required to achieve impartiality. To be frugal: all that the POPs need to be ignorant of are their roles as individuals in the world of POLs. That is to say, it would be possible to secure impartiality while allowing the POPs to know the entire history of the world—not only the general conditions governing it, but the actual course of history, and indeed the alternative courses of history which would be the result of different actions by individuals in it, and in particular to know that there would be in the world individuals *a*, *b*, . . . *n* who would be affected in specific ways by these actions—*provided* that each of the POPs did not know which individual he was (i.e., whether he was *a* or *b*, etc.). Impartiality would be secured even by this very economical veil, because if a POP does not know whether he is *a* or *b*, he has, however selfish, no motive for choosing his principles so as to suit the interests of *a* rather than those of *b* when these interests are in conflict.

A superficial reading of Rawls' rather ambiguous language at 137/4, 12/12 and 198/20 might lead one to suppose that this "economical veil" is what he has in mind. But this cannot be right, in view of 200/17 and other passages. We need to ask, therefore, why Rawls is not content with it, if it suffices to secure impartiality. The answer might just be that he is un-

<sup>9</sup>"Rules of War and Moral Reasoning", *Philosophy and Public Affairs* 1, 1971/2, p. 166.

<sup>9</sup>Rawls' word; I have commented on his use, and given reasons for preferring the word 'universal', which *he* uses for something else, in my paper "Principles", *P.A.S.* 73 (1972/3), p. 2.

<sup>10</sup>I have hinted why in "Principles" (*op. cit.*), p. 4. For my own answer to the "rigging" difficulty, see my *Freedom and Reason*, p. 107.

clear as between two things : (1) the POPs' not knowing which of them is going to be *a* and which *b* ; (2) their knowing this, but not knowing how *a* and *b* are going to fare. Much of his language could bear either interpretation. And 141/25 seems to imply that Rawls thinks that the "economical veil" would allow the POPs to use threats against each other based on the power which as individual POLs they would have ; but this is obviously not so if they do not know *which* individuals they are going to be, however many particular facts about individuals they may know.

Nevertheless, sooner than accuse Rawls of a mere muddle, let us look for other explanations. One is, that he wants, not merely to secure impartiality, but to avoid an interpretation which would have normative consequences which he is committed to abjuring. With the "economical veil", the rational contractor theory is practically equivalent in its normative consequences to the ideal observer theory and to my own theory (see above and below), and these normative consequences are of a utilitarian sort. Therefore Rawls may have reasoned that, since an "economical veil" would make him into a utilitarian, he had better buy a more expensive one. We can, indeed, easily sympathize with the predicament of one who, having been working for the best part of his career on the construction of "a viable alternative to the utilitarian tradition" (150/12), discovered that the type of theory he had embraced, in its simplest and most natural form, led direct to a kind of utilitarianism. It must in fairness be said, however, that Rawls does not regard this motive as disreputable ; for he is not against tailoring his theory to suit the conclusions he wants to reach (see above, and 141/23, where he says, "We want to define the original position so that we get the desired solution"). I shall be examining in the second part of this review the question of whether Rawls' thicker veil *does* help him to avoid utilitarianism ; it is fairly clear from §28 that he *thinks* it does.

A further motive for the thicker veil is a desire for simplicity both in the reasoning and in the principles resulting from it (140/31 ; 142/8 ; but cf. 141/22). By letting the POPs know only the general facts about the world in which the POLs live, and also by other devices (e.g., 95/14, 96/6, 98/28), Rawls effectively prevents them from going into much detail about the facts. This means that his principles can and must be simple ; but at the same time it raises the question of whether they can be adequate to the complexities of the actual world. Rawls is, in fact, faced with a dilemma. If he sticks to the "economical veil", then there will be no difficulty of principle in doing justice even in highly specific and unusual cases in the actual world ; but this will involve very complex calculations, in advance, on the part of his POPs. On the other hand, if, in order to avoid these complex calculations, he limits the POPs' knowledge to "general" facts about the world, he is in danger of having his POPs choose principles which may, in particular cases, result in flagrant injustice, because the facts of these cases are peculiar.

This is merely the analogue, in Rawls' system, of the dilemma which afflicts utilitarians, and which I have tried to solve in two articles already referred to.<sup>11</sup> The solution lies in distinguishing between two levels of moral thinking, in one of which (for use "in a cool hour") we are allowed to go into all the details, and in the other of which (for normal use under conditions of ignorance of the future, stress and temptation, and in moral education and self-education) we stick to firm and simple principles which are most likely in general to lead to right action—they are not, however, to be confused with "rules of thumb", a term whose indiscriminating use has misled many. The first kind of thinking (let us call it "level-2") is used in order to select the principles to be adhered to in the second kind ("level-1"), choosing those principles which are best for situations likely to be actually encountered. If this kind of solution were applied to Rawls' system, he would allow his POPs to know everything but their individual roles as POLs (the "economical veil"); but since their task would be to choose the best *level-1* principles for the thinking of POLs, they could still, since these principles have to be simple and observed only in general, attend only to the general facts about the POL society and the general run of cases. The contract would then not be a contract to act universally in certain ways, but rather a contract to employ certain firm principles in the moral education of POLs themselves and their children, and to uphold such principles as the norm in their society. For unusual cases, and for those in which the principles conflicted, the POLs would be allowed (in Aristotelian fashion)<sup>12</sup> to do a bit of POP-thinking for themselves.

Rawls does not adopt this solution, although he shows some awareness of the distinction between level-1 and level-2 thinking on 28/19. Ross's different but related distinction between "*prima facie* duties" and "duties all things considered" is referred to and indeed used on 340/15. On the whole Rawls' principles are treated as unbreakable ones for universal observance (e.g., 115/36); but they are supposed to have the simplicity which in fact only level-1 principles can, or need, have (132/17). Other passages which *might* be relevant to this question are 157/32, 159/16 ff., 161/17, 304/13, 337/11, 340/28, 341/14, 454/6; but I have been unable to divine exactly what Rawls' view is.

He has tried to get over the difficulty of conflicts between principles and unusual cases in two ways. The first is by means of a rigid "lexical" ordering of his principles (which could be guaranteed in unusual cases to yield absurd results); the second is by his "four-stage sequence" (195 ff.), whereby the "veil of ignorance" is progressively lifted, and at each "lift" the knowledge of extra facts is absorbed and the principles expanded to deal with them. The sequence ends with the complete disappearance of the veil. Since Rawls can say this, he cannot have any objection on grounds

<sup>11</sup> "Principles" (*op. cit.*), pp. 7 ff.; *Philosophy and Public Affairs* 1 (1971/2), p. 166.

<sup>12</sup> *Eth. Nic.* 1137 b 24 ff. (cf. Rawls 19/9, 138/20).

of practicability to unrestricted knowledge from the start, and his reasons for forbidding it must be theoretical ones.

The four-stage sequence would only work if at each later stage the principles inherited from the stage before *determined*, in the light of the new information, what further principles were to be adopted. At least this is so, if the method is required to be rigorous; Rawls can perhaps escape this requirement by using intuitions all down the line. But if the principles chosen in the original position do *determine*, in conjunction with each new batch of facts, all the additional principles that are to be adopted at each stage, then the moral law is likely to turn out to be an ass. Some victim of the application of one of these lower-order principles may be found complaining that if the POPs had only *known* about him and his situation, likes and dislikes, then they would have complicated their principles a little to allow them to do justice to him (perhaps he does not give a fig for the "priority of liberty"; or perhaps his preferences do not coincide with the POPs' ranking of the "primary goods"). If it were rigidly applied, Rawls' system would be like a constitution having a legislature in which reading of the newspapers was forbidden, and law-courts without any judicial discretion. But of course he does *not* apply it rigidly.

I will conclude this part of the review by showing why the ideal observer theory, the rational contractor theory and my own theory must, on certain interpretations of each of them, yield the same results. As pointed out above, the "economical veil" version of the rational contractor theory secures impartiality between the individuals in society. The ideal observer theory includes impartiality as an express stipulation. My own theory secures impartiality by a combination of the requirement that moral judgments be universalizable and the requirement to prescribe for hypothetical reversed-role situations as if they were actual (I am not sure whether the second is an independent condition or not). So, as regards impartiality, the theories are on all fours. Next, some degree of benevolence is required by all three theories; the ideal observer is expressly required to be impartially benevolent; my universal prescriber, since he has to treat everybody as one and nobody as more than one, and since one of the persons included in "everybody" is himself, to whom he is benevolent, has to be positively and equally benevolent to everybody; the rational contractor, although he is selfish, does not know which individual POL it is whose interests he should favour (since he does not know which is himself) and so his selfish or partial benevolence has the same results as impartial benevolence. For the same kind of reason the ideal observer and the universal prescriber, though they have additional knowledge (*viz.*, knowledge of their own individual roles, if any, in the situations for which they are prescribing) are prevented by the previous requirements from using it for selfish ends.

Rawls himself says that the ideal observer theory leads to utilitarianism

(185/24) ; and—at least if it takes a certain form, if it involves what Rawls calls 'sympathetic identification' with all affected parties—this seems plausible. In stating this form of the theory, he echoes some phrases of my own,<sup>13</sup> and later treats my theory and the ideal observer theory as equivalent.<sup>14</sup> So, then, the rational contractor theory, in the version I have been discussing (which is not Rawls') should also lead to utilitarianism. Rawls is aware of this possibility (121/33). He even seems to imply on 149/3 that *his own* theory is practically equivalent to the ideal observer theory ; but this is not his usual view. I shall leave to the second part of this review the difficult task of deciding whether Rawls, by the departures he makes from this simple version of the rational contractor theory, succeeds in establishing a non-utilitarian conclusion.

Of the three theories that I have just shown to be practically equivalent, it is largely a matter of taste which one adopts. Philosophers will differ in the use they like to make of dramatizations of their theories, and in the particular scenarios chosen. Such dramatizations do not help the argument, though they may help to expound it ; Rawls himself seems to agree (138/31). For myself, I think such devices useful (though they can also mislead), and I had much greater hopes of Rawls' enterprise than have in fact been realized. For, knowing that the simplest and most natural version of the rational contractor theory was practically equivalent to my own position, I was optimistic enough to hope that Rawls' elaborate exploration of the normative consequences of such a theory might illuminate those of my own, and thus enable me (in the most favourable outcome) simply to plug in to his results. Such good luck, however, seldom befalls philosophers ; and in fact Rawls' constant appeal to intuition instead of argument, and his tailoring of his theory to suit his anti-utilitarian preconceptions, have deprived it of the value which it could have had as a tracing of the normative consequences of views about the logic of the moral concepts.

It is interesting that in his peroration (587) Rawls as good as drops into an ideal-observer way of speaking. I myself should be happy to use any of these images (including C. I. Lewis's "all lives *seriatim*" picture<sup>15</sup>). But the work needs to be done on the logic of the argument, which has to be shown to be valid by the procedures of philosophical logic, involving the analysis of concepts, natural, or if need be artificial. Without this, a "theory of justice" is nothing but a suggestive picture.

*Corpus Christi College, Oxford*

<sup>13</sup>186/30, 34 ; *Freedom and Reason*, p. 123.

<sup>14</sup>See *Freedom and Reason*, p. 94 n.

<sup>15</sup>*Analysis of Knowledge and Valuation*, p. 547 ; Rawls 189/12 ; cf. my *Freedom and Reason*, p. 199, where through ignorance I failed to acknowledge Lewis's use of this picture.