Ernst Tugendhat

Comments on some Methodological Aspects of Rawls' "Theory of Justice"*)

Abstract: In the first part of the paper Rawls' conception of a „reflective equilibrium" with our „considered moral judgements" is criticized. Moral judgements cannot form a court of appeal for the justification of moral principles, since they are themselves in need of justification. An analysis of the meaning of the sentences in which moral judgements are expressed is called for in order to establish their method of justification.

In the second part of the paper the consequence which Rawls' repudiation of semantic analysis has had for his conception of the „original position" is discussed. In retrogressive extension of his four-stage-sequence a zero-stage is postulated which represents the moral point of view. At this stage the reasons would have to be given for adopting the original position and for conceiving it with just those characteristics that Rawls has assumed. Only thus can the advantanges and disadvantages of these characteristics be analytically assessed.

When we compare Rawls' Theory of Justice with the two most important traditional modern ethical theories, the utilitarian on the one hand and the Kantian on the other, a curious contrast emerges concerning the content and the method of Rawls' theory. Rawls' theory is directed against utilitarianism in its content, and in this respect Rawls is and considers himself to be close to Kant's conception. In his methodological convictions, on the other hand, Rawls opposes a conception that is based on "the analysis of moral concepts and the apriori" and believes that the object of a moral theory is to give a theory of our "moral sentiments" (Rawls 1971, 51). Moral theory is to be checked against a class of *facts,* "our considered judgments in reflective equilibrium" (51). Rawls claims that this "is the conception of the subject of the classical writers at least down through Sidgwick". However, the tradition of classical writers to which Rawls aligns himself here is primarily the utilitarian tradition. His reference to Aristotle in a note to the sentence just quoted is disbutable. Hare, who, though not mentioned, appears to be the main target of this and similar passages, is by no means the first philosopher who built his ethical theory on an analysis of the meaning of "good" and other conceptual analyses; the same was true of Kant and, even if in a very different way, of Aristotle.

A possible explanation of the proximity in method to his foremost opponent — utilitarianism — is that Rawls himself primarily belongs to the utilitarian tradition. The argument between the Theory of Justice and utilitarianism appears to be an

*) Unrevised version of a lecture given at Bad Homburg in April 1976 within a symposion about (and together with) Rawls organized by G. Patzig.

Analyse & Kritik 1 (1979), Heft 1, S. 77-89

argument between kinsmen. But Rawls' theory is close to Kant's not only in its content; to justify his theory Rawls makes use of a distinctly formal device, a contract theory, of which one may wonder whether it is not a heterogeneous element within his other methodological assumptions. Now it is true that Rawls thinks that it is just this device which enables him "to leave questions of meaning and definition aside and to get on with the task of developing a substantive theory of justice" (579). And it must also be admitted that he admirably connects this contractarian approach with the methodological conception of the reflective equilibrium. The true principles of justice are according to Rawls those that would be chosen in what is being described as the "original position", but to give the original position this significance is itself justified only "if the principles which would be chosen match our considered convictions of justice or extend them in an acceptable way" (19).

However, the precise significance and the justificatory force of the original position for Rawls' conclusions is what has caused critical readers of *The Theory of Justice* the greatest difficulties. Thus the conjecture appears not too bold that the disregard for conceptual analysis which for Rawls resulted from his conception of reflective equilibrium had a harmful effect on the clarity of what precisely is intended or attained with the notion of the original position. I shall therefore proceed in this paper in two steps. First I shall comment on and express my doubts about Rawls' conception of a moral theory and the concept of reflective equilibrium. In the second part of my paper I shall then deal with the effects which Rawls' dismissal of conceptual and analytical considerations had on his conception of the original position as a justificatory device.

I

The task of moral philosophy is according to Rawls to find principles which fit our "considered moral judgments". He adds that to put it in this way is only a first approximation, since a person is likely to change some of his considered moral judgments in the light of principles and especially in the light of various proposed principles. So a reciprocal adjustment of considered judgments and principles takes place, and when this process comes to a provisional standstill, Rawls speaks of a "reflective equilibrium" (20, 48).

It is by no means easy to understand this conception. Rawls explains that one must regard "moral theory just as any other theory" (578). This seems to presuppose that all theories are basically similar. Rawls mentions linguistics (47), physics (49), mathematics (51) and the philosophical theory of the justification of deductive and inductive inference (20). Now in each of these cases the relationship between principles and facts is significantly different from every other case. A linguistic theory has a subject-matter – the competent native speaker – that is itself guided by principles or rules, while in the case of a science like physics the data themselves have nothing to do with principles; the principles are merely in the theory. In this respect what Rawls seems to have in mind for moral theory is at least closer to linguistics than to physics. But even in the case of linguistics it would not make very much sense to speak of a reflective equilibrium in the sense just explained. Rawls admits that "we may not expect a substantial revision of our sense of correct grammar in view of a linguistic theory" (49). But this difference between linguistic and moral theory does not appear to be as contingent as Rawls makes it out to be. This difference must have something to do with the fact that moral theory, as Rawls describes it, is carried through in the 1st and 2nd person. This is what Rawls calls "the Socratic aspect" of moral theory (49, 578). It is obvious that only if the data belong to the same person who is doing the theory can the data change in the light of principles which the theory puts forward.

Of course one can carry through a theory of the moral sentiments very similar to the one described by Rawls except that it would be a 3rd person theory. In this case the theory that would emerge would simply lack the aspect of reflective equilibrium, the data would not change in the light of the principles, and we would have a straightforward empirical theory. Any psychological or anthropological theory of the sense of justice of a group or society would be of this type. The comparison Rawls makes with linguistic theory would thus seem to fit this other type of a moral theory which is not the one pursued by Rawls.

Now if this is so, we should be able to throw further light on Rawls' conception of a moral theory by asking why in the case of a moral theory the theory in the 1st and 2nd person may be significantly different in its structure from the theory in the 3rd person whereas this is not so in the case of linguistics. Why do we not have a linguistics that is "Socratic"? And why do we not have a special motive to do linguistics in the 1st person but may have a special motive for a moral theory in the 1st person? A first answer seems obvious: our linguistic competence is not improved by reflection on its principles while our sense of justice may be improved by such a reflection. This explains an aspect of Rawls' conception of moral theory that I have not yet mentioned. He says: "a conception of justice ... is a matter ... of everything fitting together into one coherent view" (21). This coherence theory of moral justification is obviously a corrolary of the conception of reflective equilibrium. We would not speak of a coherence theory e.g. in linguistics nor in any other empirical theory, because in every such theory the principles have to agree with the data and there is no question of a reciprocal readjustment.

All of this remains, however, still too much on the surface. It seems true that the fact that we can have a significantly different 1st person theory in ethics but not in linguistics is connected with the fact that reflection can improve our moral sense, but not our linguistic competence, and it also seems true that this improvement has something to do with increasing coherence, but the question remains what the reason is for these connections.

However, this is as far as I could get by way of a mere elucidation of Rawls' conception. Although I had to introduce some distinctions which Rawls himself does not make and although these distinctions seem to me to show that Rawls' claim that one can regard moral theory just as any other theory is not true even for his own conception of moral theory, all of this was meant only to help us to understand Rawls' own conception, and I would hope that he could agree with me so far. But if he should do so, it seems difficult to avoid a further step which can no longer be understood as a mere clarification of Rawls' view but would show that, if duly clarified, this view gives way to another conception.

So I now return once again to the difference between a moral theory and a linguistic theory. The most obvious difference between the subject matters of the two theories has not yet been mentioned. This difference is implied in Rawls' use of the expression "considered *judgments*". What Rawls calls the "facts" with which moral theory has to deal are a certain kind of beliefs, beliefs about what is right or just. The discursive character of these facts is obscured when one speaks of moral *sentiments.* Now beliefs or, to use Rawls' expression, judgments have the peculiarity that they are connected with a truth claim or, if this seems preferable, a claim of validity. The standard linguistic expression of a belief or judgment is what is called an assertoric sentence, and it is the defining character of such sentences that they can be true or false. It is of course controversial whether value judgments or normative judgments can be "truly" true or false. But it cannot be controversial that they are, if I may say so, "phenomenologically" true or false. To restrict myself to Rawls' favourite word, "just", it is obvious that sentences which express a judgment or belief that so and so is just or unjust have all the characteristics of any other assertoric discourse. We use, when we express what we believe to be just or unjust, such adverbs as "really", "truly", "apparently", "seemingly"; we say such things as: "I used to believe this was just, I then doubted whether it really was, and I now know that it is not" etc.

Now this fact that what Rawls calls our sense of justice consists in a belief system contains, I think, the explanation why there is a significant difference between a moral theory in the 1st and 2nd person and a moral theory in the 3rd person, but this difference now proves to be much deeper than what could appear on the basis of Rawls' own account. The important point is that what Rawls calls the facts for the theory are in this case facts connected with a truth claim. For the persons whose judgments they are they are not just facts against which a theory can be checked, but for them they are, being beliefs, themselves items that are susceptible of being checked. One aspect of any belief system is, of course, that it must be coherent; if it is self-contradictory, it cannot be held. But this alone cannot explain the importance that the reflection on principles apparently has for 1st person moral theory. We must distinguish between different kinds of belief systems. Believes about matters of fact are characteristically justified, directly or indirectly, by observation. Moral judgments, on the other hand, if they can be justified at all – and they at least pretend to be justifiable – can only be justified by principles. The reason then why principles become so important in morals from the point of view of the persons themselves who make the moral judgments is that they apparently play a central role in the process of justification. Thus it seems that Rawls, if he thinks of 1st person moral theory, has put the cart before the horse. It is not the principles that are to be checked against the particular moral judgments but the other way around. Lest I be misunderstood I hasten to add that of course even now the possibility for a 3rd person moral theory is open, and this is like any other empirical theory a theory in which it is the conjectured principles that are to be checked against the particular judgments of those persons whose sense of justice is being studied. But of course such a theory is a theory not of what is just but of what the persons that are being studied believe to be just.

Perhaps I was too dogmatic in asserting that moral judgments can be justified only by principles. What I claim is only that, if we make a 1st person moral theory at all, we must realize that our moral judgments are items which, according to their own sense, do not form a court of appeal but are in need of a court of appeal. The primary question for anybody who starts to reflect on his moral judgments is the question how this sort of judgments can be justified. Rawls has managed, by his unwarranted assimilation of 1st person moral theory to 3rd person moral theory, to blind himself against this question. But then his attacks on meaning analyses are without weight. If the problem of justification does not exist, we indeed are in no need of a means of getting to grips with it. But if it does exist, one may well wonder how the question of the method of justification of a kind of sentences is to be tackled without an analysis of the meaning of these sentences.

There remains one illuminating reference that Rawls makes in connection with his concept of reflective equilibrium that I have not yet mentioned. It is the reference to Nelson Goodman's apparently similar theory concerning deductive and inductive inference (20, note). According to Goodman "principles of deductive inference are justified by their conformity with accepted deductive practice" and "rules and particular inferences alike are justified by being brought into agreement with each other" (Goodman 1965, 63, 64). This reads indeed very similar to Rawls' conception of reflective equilibrium. Now Goodman's view is itself not uncontroversial, but anyway there is a notable difference between his view and Rawls'. The "facts", to use Rawls' expression, consist in Goodman's case not in judgments, but in procedures used in justifying judgments. To transfer Goodman's idea to the case of moral theory would thus lead to a different program from the one that is being advocated by Rawls. The program would now aim not at the justification of moral principles, but at the justification of the methods of justification. It would consist in the analysis of the rules of valid moral argument. I can leave it open whether the best we can do in this question is, in analogy to what Goodman says, to justify the principles of valid moral argument by testing them against the accepted practices of moral argument; at any rate we would be much less sure in the case of moral argument what we should count as "accepted practices" than in the case of deductive and inductive inference. The alternative to such a conception would be a conception such as Hare's: that the rules of valid moral argument follow from the logical structure of these sentences. In the present context I can leave this issue open, because my quarrel with Rawls is not that I don't agree with an answer he gives but that he does not even pose the question and places a coherence theory in its stead. It is of course possible to doubt that moral judgments can be justified at all; it is possible to maintain that their truth-claim only gives them an appearance of being justifiable and that there are no decision procedures to back this claim. But this contention could in turn only be founded on an analysis of the logical structure of these sentences. Rawls takes neither a positive nor a negative position toward this question, but simply brushes it aside.

Before closing this part of my paper, let me make a step toward reconciliation. It would be a misunderstanding to think that the result of what I said would be that the notion of reflective equilibrium must be abandoned. It would only have to be interpreted differently. The considered moral judgments are indeed the point of departure for any moral reflection, but their value is heuristic, not that of a court of appeal.

Kant, for example, started out, in the 1st section of the *Grundlegung,* with an analysis of our "considered moral judgments"; the result of this analysis he then checked in the 2nd section by an analysis of the concept of an unconditionally good action. Thus the checking proceeded in the contrary direction to the one advocated by Rawls. The 2nd section was for Kant the decisive one, and for us, who no longer share all the "considered moral judgments" of Kant's time, it is that part of his moral theory that has remained valuable. What Kant accomplishes in the 2nd section of the *Grundlegung* can also show how unsubstantiated Rawls' claim is that questions of meaning and definition are useless for settling substantive moral problems. This is an argument *ad hominem,* since Rawls in § 40 accepts Kant's substantive conclusions without caring for their formal derivation in Kant.

II

It might seem that Rawls' contractarian theory provides us in practice with what in theory he appears to deny: a method of justification. But of course there is no inconsistency here, because Rawls can easily incorporate his contract theory into his doctrine of reflective equilibrium. He in fact maintains that this procedure of justification can itself only be justified by showing that its output tallies with our considered moral judgments. Thus Rawls' contract theory is to a certain degree neutral in relation to the controversy with which I dealt in the 1st part of my paper. A philosopher who disagrees with Rawls' contention that rules of moral reasoning are justified if they lead to our considered moral judgments could still agree with Rawls' contract theory as an adequate setting for moral argument which he then would have to justify independently.

Rawls even goes so far as to meet such a philosopher halfway. He says "that there is a broad measure of agreement that principles of justice should be chosen under certain conditions" (18), and he justifies his conception of the original position by trying to show not only that the principles chosen in this position tally with our considered moral judgments but that this position also satisfies those conditions which are generally thought characteristic for the "moral point of view" (120). Rawls seems to give these conditions that are characteristic for the moral point of view a similar status that he gives to our considered moral judgments. Although he is, as far as I can see, not very specific on this point, I presume he would say that we not only have considered moral judgments about particular moral questions but also have considered judgments about the conditions of moral reasoning, and a valid theory should arrive at a reflective equilibrium with both sides.

Now these conditions of moral reasoning have obviously an abstract and – *pace* Rawls – conceptual character; they belong into that line of clarification which a philosopher would follow who would want to inquire whether Rawls' description of the initial situation can be taken as the adequate position for valid moral arguments. Of course such a philosopher would also want to follow this line further back than Rawls, perhaps up to a point where the matter could be decided by a conceptual or logical analysis. But I shall not try to do this. I wish to carry these reflections on the conditions of moral argument not further back than Rawls himself, or at least not much further back, because I do not intend a criticism from the outside.

What I want to discuss in this part of my paper is the question whether the tendential aversion to an analytical and conceptual approach which arose from Rawls' methodological conception did not have damaging effects on the way in which Rawls introduces the contractarian position. I do not want to say that these effects or even the negative attitude toward conceptual questions is a *necessary* consequence of Rawls' methodological conception. From the fact that an apriori argument on the validity of moral reasoning would have to be conceptual it does not follow that if you don't attempt such an apriori argument you don't need conceptual analyses. And for such a formal conception as Rawls' contractarian position one would have expected that a conceptual approach would have been especially important. Now I also don't want to exaggerate and do not wish to insinuate e.g. that Rawls is conceptually unclear. What I wish to maintain is that his introduction of the original position is not sufficiently analytical to be properly assessed.

It has a strongly synthetic character in being a relatively many-sided scheme, and Rawls has not explained step by step which of its aspects follow from those conditions that he assumes are generally thought characteristic for moral argument and which aspects he has introduced for other reasons; and he has done very little to show the superiority of his conception in comparison to other conceptions which would also fit those conditions of moral argument. It apparently seemed sufficient to Rawls to point out that a) many aspects of the original position do agree with those conditions and b) that the principles chosen in the original position agree with our considered moral judgments. This state of affairs must of course appear especially unsatisfactory to those of us who would like to look at Rawls' proposal as a proposal of the true condition of valid moral reasoning, but I shall show that it also has doubtful effects from Rawls' own point of view, concerning the agreement with our considered moral judgments.

A fundamental assumption of Rawls that I do not wish to dispute is that the principles of justice and moral principles in general are not something that is given to us, presumably in some intuition, but something that we arrive at actively, in an act of choice under certain conditions. These conditions for moral choice are for Rawls circumscribed by what the calls the "initial situation" and by its further specification through the "philosophically most favoured interpretation" which is then called the "original position" (121, 146). Strictly speaking only the principles for the basic structure of society are chosen under the conditions of the original position. Rawls envisages a "four-stage-sequence" of increasing concreteness of the problems that have to be decided (§ 31).

Now what is being obscured in the way Rawls introduces the original position is that this introduction represents itself an act of choice. .The original position has to be *adopted* as the *best* position from which to decide on moral principles in comparison to other possibilities such as e.g. the impartial observer theory (cf. 184 ff). When Rawls gives reasons why we should adopt the original position as the most adequate position in which principles of justice are to be chosen, he is therefore operating at a stage that is preliminary to the first of his four stages. This zero-stage of moral choice is of course not characterized by a veil of ignorance, since the veil of ignorance is one among other things that is being chosen at this stage. It is also not a hypothetical situation, since, again, the hypothetical condition of the initial situation is something that, at the zero-stage, is an object and not a condition of choice. Finally, the kind of choice that is called for at the zero-stage is not a "rational choice" in the "narrow sense, standard in economic theory" which is characteristic for the choice that is to be taken in the initial situation (14). Since the deliberation that is necessary at the zero-stage must be considered the foundation of moral philosophy, Rawls' contention that "moral philosophy" is to be conceived of as "part of the theory of rational choice" (172) is at least not true of this fundamental first step.

Now the choice that is called for at the zero-stage must also stand under certain conditions. But these cannot be determined by certain subjective conditions (like ignorance, rationality, etc.) but only by the kind of thing that is to be chosen. The thing to be chosen seems to be: an adequate representation of the moral point of view. The conditions for the zero-stage choice are therefore the defining characteristics of the moral point of view. Now these can be determined in either of two ways. They can be derived from a logical analysis of what it can mean to justify moral propositions, and, as I have said, I shall not follow this direction, since it is contrary to the one taken by Rawls. Or one simply picks up, as Rawls does, those conditions that appear to be generally accepted as characteristic for the moral point of view.

Now since Rawls has not explicitly set out what I call the zero-stage, he has not begun as we should have expected him to begin, by a full enumeration of these conditions. Therefore he has left it unclear which aspects of the original position follow from these conditions and which of its aspects he has chosen for other reasons. This unclarity can only be removed by assembling the relevant things which Rawls says in diverse places. The nearest that Rawls comes to an enumeration of such conditions is the enumeration of "the formal constraints of the concept of right" in § 23. The most important of these "formal constraints" are "generality" and "universality". As Rawls understands these principles, they do not yet seem to imply impartiality. But this notion is rightly stressed by Rawls at several places as fundamental for the moral point of view (cf., in connection with the introduction of the original position, pp. 12 and 18). I feel less sure about how much weight he gives to the condition of "autonomy" (the principles are to be "self-imposed", 13). It is one of the virtues of the original position that it satisfies this condition, and this condition does not seem to be met by the impartial spectator theory, and yet, where Rawls discusses this theory (§ 30), he does not criticize it on this account. If we don't include the condition of autonomy, the moral point of view might be summarily characterized as the point of view at which such principles of action would be chosen that are in the interest of everybody. The condition of autonomy is included, if we reformulate this by characterizing the moral point of view as a condition of choice according to which only such principles are chosen to which everybody could agree.

These characterizations are extremely rough and would need further elaboration. The important point is that to characterize the zero-stage a mere enumeration of several conditions is insufficient; we have to define the moral point of view by some such comprehensive characterization. In contrast to the diverse hypothetical models such as the contract model or the ideal observer model the moral point of view does not represent a hypothetical choice situation but the situation of moral choice within our real life. (It is true that even this choice contains hypothetical elements, when I say, e.g., that such principles are chosen to which everybody *could* agree, but the choice itself is not hypothetical.) It should not be controversial that moral philosophy cannot *begin* with a hypothetical situation but only with the moral point of view as a phenomenon of our actual life.

It would now be the second step to show that within this zero-stage we have reasons to adopt a hypothetical position that is to serve as a representative for the moral point of view. This Rawls has omitted to do. What he has shown was merely that the original position incorporates the *same* conditions that are characteristic for the moral point of view. He has not explicitly shown why it is *preferable* to shift the choice situation from the zero-stage to the original position. Thus it has remained unclear whether the reasons for this shift are a) reasons that improve the moral perspective itself or b) reasons of practicality or c) reasons that have something to do with the special subject matter of the choice of the principles for the basic structure of society, but perhaps not for other moral choices. And, of course, Rawls has made no attempt to weigh the advantages of his proposal against its conceivable disadvantages.

In the remainder of my paper I shall only sketch answers to these questions. The most characteristic difference of the contractarian model from the original moral point of view is that it allows to separate the act of choice from the regard to the interests of everybody; impartiality is attained not by the intention to come to an agreement or by some other intentional effort contemporaneous with the act of choice, but by the previous application of a veil of ignorance, with the result that the choice can now be a "rational choice" which aims only at one's own advantage, and with the further result that to speak of an "agreement" is really redundant, since the agreement would be "unanimous" (139). ("Therefore, we can view the choice in the original position from the standpoint of one person selected at random." (139))

It seems that the main reason why Rawls considers the original position preferable to the original moral point of view is that it allows to conceive of the theory of justice as "part of the theory of rational choice" (16), which appears to be something more manageable than the Rational choice with a capital "R" of which we would have to speak at the zero-stage. However, it would remain to be tested whether in practice a rational choice, when carried through in the original position, really leads to results that are in some way better than those to which we would be led at the zero-stage. The main test case is here obviously Rawls' justification of the "difference principle" by this method, and I shall come back to this problem. At any rate, the advantage of being able to apply the theory of rational choice would be an advantage of practicality; it would improve not our concept of what is just but the decision procedures to arrive at just results. This, of course, would indeed be an advantage which we should not underestimate.

Another feature in favour of the original position which Rawls often mentions is that it allows the application of the notion of "pure procedural justice" (136, 304). "Pure procedural justice obtains when there is no independent criterion for the right result: instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is" (86). As far as I can see, pure procedural justice is necessary only when no more direct decision procedures are available. It is therefore not suitable for a clarification of our general notion of justice, but is a limited though important moral device, adequate for the decision of certain political and moral problems and not of others. If the applicability of this notion were a prerogative of the original position, we would have here another major advantage of practicality and besides one that is restricted to certain subject-matters. The reason why the original position appears especially apt to allow for pure procedural justice is that this type of justice implies a preliminary agreement to follow certain rules. But there is no reason why such agreements, whether hypothetical or real, cannot be arrived at directly and *ad hoc* from the original moral point of view. The conception of the initial situation bases the whole of morality on a preliminary hypothetical agreement. The original moral point of view does not conceive of the concept of right in general in this way, but leaves it open to determine it in this way in those cases in which this is called for.

I now turn to the problem of the "veil of ignorance". It seems that here in particular Rawls has conflated several different aspects. He introduces the veil of ignorance in his usual synthetic fashion in one grand stroke without explaining for which reasons the several parts of this veil are necessary. The only justification that he gives for the veil in its entirety (12, 136) can really serve as a justification only for a part of it: to insure impartiality it would have been enough that in the initial situation everybody be ignorant of his own identity (cf. Hare 1973, 89f). Rawls assumes in addition that in the initial situation everybody must also be ignorant of "the particular circumstances of their own society" (137). One reason for this assumption is that "questions of social justice arise between generations as well as within them" (137) but this alone would not be a sufficient reason to require that in the initial situation even "the course of history is closed" to us (200). These further limitations do not arise from the requirement of impartiality but because "without these limitations on knowledge the bargaining problem of the original position would be hopelessly complicated" (140). Here then we have another special feature of the original position that does not correspond to the moral point of view as such but is added for reasons of practicality. And again these limitations seem to be appropriate only for certain moral problems, although these may be the most fundamental ones: in the 4-stage-model this part of the veil is gradually lifted (§ 31).

Surprisingly Rawls says that in the last stage – "the application of rules to particular cases" – the veil of ignorance is removed entirely (199). This must be a mistake, if the type of choice in this last stage is still to be of the self-interested kind and if the outcome is to be nevertheless impartial. Only that part of the veil can be entirely removed at the 4th stage which has been added for reasons of practicality.

Again it may be asked whether the additional veil of ignorance which does not arise from the requirement of impartiality is really an asset of the original position. Since it amounts simply to a decision to disregard all those facts that appear irrelevant for the solution of a problem, this can of course be carried through just as well directly from the moral point of view, and probably better, because at the moral point of view we would not once and for all be cut off from all information; the question *which* facts are irrelevant for the choice e.g. of the basic principles of justice could remain an open question during the process of deliberation. In Rawls, on the other hand, there seems to be a tendency to ignore from the beginning all such aspects of social life the comparative value of which is not quantifiable. Here it is the method of rational choice which seems to be responsible for the veil of ignorance.

Now whether all these additional features which distinguish the original position from the original moral point of view are really advantages or not: *if* they are advantages, they are advantages in practicality, not morality; they are introduced to make the decisions more manageable. This was to be expected, because it would be contradictory to think that the strictly moral features of the moral point of view could be improved by changing its conditions of choice, since what we mean by "moral" or "right" is defined by the moral point of view. Rawls' suggestion that we could "define the concept of right by saying that something is right if and only if it satisfies the principles which would be chosen in the original position to apply to things of its kind" (184, cf. also 111) is a *petitio principii* and obscures the fact that the moral adequacy of the original position must be assessed from the point of view of the zero-stage. The concept of right can only be defined by saying that something is right if and only if it is the outcome of a decision procedure which begins at the zero stage (which of course I have not adequately defined). The best that could be said for the original position would be that Rawls' 4-stage-model is the only or the most manageable decision procedure for the moral problems that arise at the zero stage. And in this case the conception of justice as fairness would indeed be vindicated.

But the question must now be faced whether the advantages in practicality which are possibly gained by the shift from the zero stage to the original position are not paid for by a loss in moral substance. It is not self-evident that if we take apart the idea of a moral (impartial) decision into the two components of a self-interested decision plus ignorance of one's identity, the result remains the same. The contractarian conception with its insistence on an (even if only hypothetical) *initial* agreement introduces into the problems of justice an element of time-lag which is not contained in the original moral point of view and in our ordinary conception of justice. It is this time lag that allows Rawls to apply the theory of rational choice, but the probability problems that enter here with their specific psychological counterpart – the expectation of chances and the disposition toward risks – do not seem to have anything corresponding in a normal moral judgment, except where by the nature of the case we cannot but adopt methods of pure procedural justice. There is of course a long way from falling back on pure procedural justice where we cannot do any better to claiming that the entire problem of social justice is a problem of justice as fairness.

Rawls thought that the transposition of the problem of justice into a problem of rational choice gave him the decisive weapon against utilitarianism. But several critics have pointed out that it appears to be a mistake that the most rational thing to do in the initial situation would be to apply the maximin principle and thus opt for an egalitarian society (cf. Lyons 1975, Hare 1973, Barber 1975). If these critics are right, what would follow from Rawls' premises would be the utilitarian conception. Would that prove that utilitarianism is right and egalitarianism wrong? Surely not, since the moral point of view clearly favours egalitarianism. Thus what seems to follow is rather that the original position is not an adequate model of the moral point of view. Suppose somebody says: "In the original position I would opt for a social system ruled by the principle of utility, because this would maximize my chances; but morally I reject such a system as unjust."

According to Rawls it would be self-contradictory to say such a thing, but it does not appear to be self-contradictory and may even be true.

I believe that there are other consequences of the transposition of the original moral choice situation into a self-interested choice situation which show that something of the moral substance gets lost. One of these concerns the argument for equal liberty of conscience. In the original position one can hardly argue as directly as Rawls for the importance of this right (206). Why should people who are only self-interested appreciate such a thing as moral conscience at all? If, on the other hand, we argue from the original moral point of view as I characterized it, we begin by considering everybody as a moral person, as a subject and not only object of moral deliberation.

The last two arguments were in part arguments *ad hominem:* they would show that the original position leads to results that do not agree even with Rawls' "considered moral judgments". However, the intention of these arguments as well as of the previous ones was not to disparage the original position but to plead for an analytical evaluation of this conception.

Bibliography

Barber, B. (1975), *Justifying Justice: Problems of Psychology, Politics and Measurement in Rawls,* in: Daniels (Ed.) (1975, 292–318)

Daniels, N. (ed.) (1975), *Reading Rawls, Critical Studies on Rawls' "A Theory of Justice",* Oxford 1975

Goodman, N. (1965), *Fact, Fiction, and Forecast,* Indianapolis N. Y. 1965^2

Hare, R. (1973), *Rawls' Theory of Justice,* in: Daniels (ed.) (1975, 81–107)

Lyons, D. (1975), *Nature and Soundness of the Contract and Coherence Arguments,* in: Daniels (ed.) (1975, 141–167)

Rawls, J. (1971), *A Theory of Justice,* Cambridge/Mass. 1971