SOCIAL CONTRACT, INTUITION AND THE PRINCIPLES OF JUSTICE*

by

WOJCIECH SADURSKI

The essential question raised about social contract theories is, of course: why should *we* be bound by a contract that did not take place, that is a hypothetical speculation and that was 'made' by persons in circumstances that do not resemble at all any of the real life situations to which this contract is applicable? In other words: why should people not in the 'original position' (to use Rawls' term for the pre-contractual stage) consider important a contract made by people in the original position? Why should *we* be bound by a contract made by *them*?

Rawls' explicit answer is less plausible than one implied in the general spirit of his theory. His explicit answer, given in almost the same wording at the beginning and at the very end of his book, is that 'the conditions embodied in the description of the original position are ones that we do in fact accept' or, if we do not, 'then perhaps we can be persuaded to do so by philosophical reflection' (p.21, cf. p.587).[1] But it is strikingly inadequate as an answer to the question about the reasons for the relevance of a hypothetical social contract to real people. Firstly, what matters is an acceptance of the *principles* reached in the contract, not of *conditions* of the description of the original position. The description of the original position, once we have decided that the whole contract is but a hypothesis, is a matter of speculation: we may agree to them or not, we may agree to some of them (e.g. the condition of rationality) and reject others (e.g. the veil of ignorance) etc. without necessarily endorsing the principles agreed upon by the original contractors. Agreement about conditions for the original position does not commit us to any substantive moral or political principles of justice.

Secondly, Rawls qualified his initial strong statement about our *de facto* agreement by a reservation that 'we can be persuaded' to agree to his proposals. But that clearly undermines the whole construction of consent as a source of the importance of the principles of justice. The contract may be binding because of an *actual*, even if only tacit, consent: it cannot be made binding by hypothetical consent as a probable result of 'persuasion'. Dworkin correctly observes that 'a hypothetical contract is not simply a pale form of an actual contract; it is no contract at all'[2]. I cannot be told that I am bound by principles on the basis of a contract to which I would have adhered (but have not) were I enlightened by a wise philosopher. Perhaps I would, perhaps I would not: that is irrelevant because in fact I did not adhere to it and therefore this contract has no relevance for me.

But to follow the track suggested by Rawls in his passages about the relevance of the hypothetical contract is to pay him lip service: it leads to a search for consent that did not take place and, in addition, consent about something else (the conditions of the original position) rather than what is in question (the principles of justice). The appropriate answer to the initial question asked in this paper is that the source of the importance of the principles of justice in Rawls' theory lies not in the consent but in the reasonableness or moral quality of those principles. Principles of justice are 'binding' in a different sense than the law: they are 'binding' by the force of moral arguments that they have in their support. The relevance of the principles of justice to us, actual people, is therefore to be found in good moral reasons for their application. The justificatory force of the original position consists therefore in the creation of circumstances in which only the power of reason moulds a contract. To ask: 'why should *we* be interested in principles reached by *them*?' is to misunderstand the nature of a social contract argument: 'they' are 'ourselves' but projected into an imaginary, more rational and more moral world. The hypothetical contract argument by its very nature rejects the distinction between:

1. an agreement between the parties to the original social contract,

2. an agreement between you and me about what is just.

If, therefore, we believe that our principles of justice should be 'reasonable' (or 'moral': here these two words are used identically) we should imagine ourselves as fully rational human beings *minus* all bias stemming from our actual self-interest and then try to see in in such a situation we could reach an agreement acceptable to all. The fact that it *would* be acceptable is not an argument in favor of its validity but of its reasonableness. However, in the case of a theory of justice, that is all we need.

Although, as I have noted, Rawls himself partly contributes to the confusion between the morality of the principles and their obligatoriness (by suggesting that we actually accept the conditions of the original position), a careful reader of his book will realise that the real grounds of the obligatoriness of the principles of justice lie elsewhere, not in a contract argument. They are discussed by Rawls in that part of his book that deals with 'natural duties'. The crucial passage explaining the citizens' obligations stemming from the principles of justice is the following:

> [A] person is under an obligation to do his part as specified by the rules of an institution whenever he has voluntarily accepted the benefits of the scheme or has taken advantage of the opportunities it offers to advance his interests, provided that this institution is just or fair, that is, satisfies the two principles of justice. (pp. 342-43)

Political obligation is derived therefore from *real* facts, not from a *hypothetical* contract. Those real facts are constituted by persons' adherence to a scheme of distribution, even if it is only a tacit adherence through actually benefitting from it. In this case Rawls' theory regains coherence: the reasonableness of justice is derived from a hypothetical social contract, its obligatoriness - from the actual social facts referred to in the last quotation.

A good deal of the confusion of the reasonableness and the obligatoriness of justice can be attributed, I believe, to the fact

that traditionally various social contract theories either did not made a distinction between both (and served both aims at the same time) or were oriented toward the derivation of principles of obligation rather than of justice. The apparent similarity between traditional social contract theories (Hobbes', Rousseau's and, in particular, Locke's) and the theory of Rawls, confused several critics[3] who took Rawls' theory to be a continuation and only a modification of the former, with a similar general aim (those critics relied here partly on Rawls' declarations - p. 11). But there is an essential difference. For Locke the social contract was fundamentally the source of the legitimacy of a government (or, more generally, of a 'Commonwealth'); for Rawls the social contract is a device for demonstrating the morality of certain principles of justice. Locke's 'government by consent' idea served to explain the grounds and principles of political obligation; in his theory there is a conceptual *iunctim* between arguments about the legitimacy of government and the principles of political morality. But this parallel justification of both can be served by a contractarian theory only under a certain condition that is met by Locke's theory but not Rawls': that in a real society we can find actual traces or proof (in the form of consent) of the original social contract. But to make this view coherent, this original social 'compact' must be thought to have been made between people as they are (or: as they were in a pre-contractual state), not as they are 'constructed' by the writer for the purposes of his theory. To be sure, Locke's social contract is not really 'hypothetical' in the same sense as the Rawlsian one: it is not a speculation about something that never took place but is useful for the purposes of social theory (as is the case with Rawls and, for that matter, Rousseau) rather, just the opposite, it is an argument about the historical thesis that in all probability a contract took place about which, for obvious reasons, we have no record. In the 'Second Treatise' Locke uses several rational and historical arguments in order to prove:

> that Men are naturally free,.....that the *Governments* of the World, that were begun in Peace, had their beginning laid on that foundation, and were *made by the Consent of the People*...[4]

He also attempts to prove that people *already* born under a government, voluntarily join the Commonwealth by tacit or express consent[5]. Thus Locke's social contract is purportedly an actual contract and it is entered into by actual consent.

It is therefore useful to distinguish among the various contractarian theories those that derive their obligatoriness from social contract and those that derive the reasonableness of certain principles from it. The first type, as exemplified by Locke or Hobbes, is constrained in its description of the conditions of a contract by actual human wants and aspirations and must trace the original contract to the consent to or benefit derived from civil society by actual people. The second type of theory, as exemplified by Rawls, is not restricted by such conditions but, in turn, it has no claim as a justification of the obligatoriness of justice. But, besides this distinction, we can draw another dychotomy among contractarian theories, along different lines. We can distinguish between those theories that take parties to a social contract as they actually are (or were) and those that construct imaginary people, deprived of certain actual characteristics and enriched with others, that actual people do not possess (or do not possess to a high degree). This distinction only partly overlaps with the former. It is clear that a social contract that is a source of the obligatoriness of certain principles must be made between people as they are (or are thought to be or have been): that guarantees some connection between the original social contract and the actual facts of social life that can be shown as traces of this contract. Characteristically, the very first phrase of Rousseau's *The Social Contract* reads: 'I mean to inquire if, in the civil order, there can be any sure and legitimate rule of administration, men being taken *as they are* and laws as they might be'[6]. But the second type: a social contract demonstrating reasonableness only, can be construed either as made between people as they are or between imaginary people, 'created' especially for the occasion. The list of alternatives produced by crossing these two classifications gives, therefore, the

following three main types of social-contract arguments:

1. a social contract among people as they are (or are thought to be or to have been in a precontractual state) grounding political obligation;

2. a social contract among imaginary, ideal, people, grounding the moral reasonableness of certain principles;

3. a social contract among people as they are (or are thought to be ... etc.) grounding the moral reasonableness of certain principles.

The first and third type of argument are compatible: Locke's social contract, for instance, justifies both political obligation and the morality of just principles. Hobbes, in turn, has only the first type of argument: from the original social contract 'are derived all the *Rights* and *Facultyes* of him, or them, on whom the Soveraigne Power is conferred by the consent of the People assembled'[7] without suggesting that those arrangements can be described in terms of 'justice'. An example of a theory containing only the third type of social contract is provided by James Buchanan who constructs a complex contractarian process starting from the stage of 'natural distribution'. The major difference between his and Rawls' theory (which exemplifies argument 2) is that in Buchanan's contract '[t]here is nothing to suggest that men must enter the initial negotiating process as equals. Men enter *as they are* in some natural state...'[8]. At the stage of a contract people are *already* in a situation of inequality and propose alternatives from their respective unequal perspectives. There is therefore no reason to predict that the post-contractual situation will be very different from the pre-contractual one. This is expressly admitted by Buchanan:

> The specific distribution of rights that comes in the initial leap from anarchy is directly linked to the relative commands over goods and the relative freedom of behaviour enjoyed by the separate persons in the previously existing natural state...To the extent that such [pre-contractual] differences exist, post-contract inequality in property and in human rights must be predicted.[9]

But, the question arises, if a social contract is aimed at deriving 'just' or 'good' principles of distribution, how can this moral quality be obtained if the contract cannot but maintain and ratify the pre-contract differences? And this is, Buchanan admits, the only reasonable expectation about the outcome of the contract. The contract is at the mercy of the 'natural distribution' and the latter cannot of course claim any moral legitimacy because, according to Buchanan's own assumption, at this precontractual stage quantities of scarce goods 'fall down' in fixed proportions onto several persons.[10] The way in which those goods 'fall down' cannot have any moral justification because it happens at the precontractual stage, and the contract is the necessary initial source of the morality of distribution. In short, if the *status quo ante* was not just, how can the outcome of a contract be just if the contract must replicate that prior state?

One could also ask how unanimous agreement is conceivable at all in a hypothetical situation which does not exclude important differences (and the awareness of them) between parties to the contract? What would be the motivation for the endorsement of or adherence to a contract ratifying prior inequalities on the part of those worse-offs and have-nots? One possible answer would be that have-nots have no choice anyway: they must agree to second-best options and the normalization of their situation under contractual agreement is better than a state of uncertainty. Although the plausibility of this answer is debatable (one could argue, for instance, that in a situation of deprivation, uncertainty is to be preferred to certainty because there would be some hope, and one could argue that have-nots would choose revolutionary violence rather than ratification of their situation), one thing seems clear: it is *not* a proper ground for the derivation of principles for a just society. Consent under economic compulsion can hardly qualify as a source of the 'agreement among free men' postulated by Buchanan as a moral source of good social arrangements[11]. Those who are worse-off under the conditions of 'natural distribution' do not choose the terms of the contract they enter: they adhere to it because they have no choice.

Buchanan's own answer to the initial question about the reasons for an agreement is slightly different: he maintains that although wealth redistribution is in principle inconceivable in the contract, both haves and have-nots will find a contract advantageous. Haves - because the contract safeguards their property rights, have-nots - because under contractual provisions richer people will 'pay differentially higher shares in those goods and services provided jointly to the whole community'[12]. But here again serious doubts arise as to whether this sort of arrangement will be found advantageous by all parties: it seems very probable that some contractors would use the contract to completely reorder the pattern of property relations.[13] What is more important for the present argument, however, is that even if such an agreement were reached, it would be a *workable* arrangement but not necessarily a *just* one. The contractual sanction for non-moral distribution cannot make it just if, by force of the assumptions made, no alternative distribution would be consented to by those who are better-off and who *know* they are better-off.

In contrast, Rawls' idea of the 'veil of ignorance' and his assumption about rational and equal contractors present strong advantages from the moral point of view as conditions for the derivation of principles of justice. It is an attempt to imagine a situation in which our views on justice are not clouded by our actual, real interests determined by our place in a distribution. In Rawls' words: 'The veil of ignorance prevents us from shaping our moral view to accord with our own particular attachments and interests' (p.516). Quite intuitively, I see a good deal of sense in asking ourselves questions like this: 'Assuming that I were not a taxpayer in my country, would I consider this taxation scheme just?' Even if, actually, I am a taxpayer in the particular country, this way of constructing a hypothetical situation is conducive to unbiased answers, in which moral views are not moulded by actual interests. I thing it is psychologically true that we have a capacity to distinguish between what our interests dictate and what is just. Even if in actual life the first approach prevails over the second, the theory of justice can and should afford the luxury of forgetting about the first approach.

But here, alas, my defence of Rawls reaches its limits. It is one thing to apply the test of the 'veil of ignorance' to particular individual persons to detect their views on justice. It is another thing to claim, as Rawls does, that in a society the adoption of this perspective by all would lead to unanimous agreement about the principles of justice. In order to prove this second point, Rawls would have to prove two important assertions. He would have to prove, first, that the conditions of the original position represent only the pure model of a rational choice situation and do not presuppose any substantive moral judgements (otherwise the argument would be circular). Secondly, he would have to prove that the same conditions of the original position cannot lead to the derivation of any other set of principles. Now, as a legion of Rawls' critics have shown, both these propositions are untenable; the common strategy for refuting them is to show that it is necessary to appeal to intuition both in order to construct the original position and in order to derive principles of justice, and that this appeal to intuition conceptually precedes the derivation of principles of justice.

Interestingly, Rawls does not deny this. With a striking (and one would say: self-defeating) frankness, he admits: 'We want to define the original position so that we get the desired solution' (p.141). He also describes his aim as 'to characterise this situation [that is, the original position] so that the principles that would be chosen, whatever they turn out to be, are acceptable from a moral point of view' (p.120) and he believes that his interpretation of the original position 'best expresses the conditions that are widely thought reasonable to impose on the choice of principles' (p.121). But all that implies that he has a view about a 'desired solution', about criteria of what is 'acceptable from a moral point of view' and a knowledge of 'conditions that are widely thought reasonable from a moral point of view' that are *prior* to the original position and the derivation of principles of justice. How can he reconcile these *a priori* moral assumptions with the general purpose of his method which is 'to derive satisfactory principles from the weakest possible

assumptions' and with the methodological postulate that '[t]he premises of the theory should be simple and reasonable conditions that everyone would grant' (pp.520-21)?

If, as Rawls explicitly admits, the derivation of the principles of justice is tailored to suit some prior moral conceptions, the obvious question is: what are the justifying grounds of those *prior* moral judgements? It is certainly not the fact that they are '*widely* thought reasonable'. The justification of moral principles is not a matter of counting heads; Rawls obviously does not hold a theory that whatever people think just, is just. He expressly states that '[f]or the purposes of this book, the views of the reader and the author are the only ones that count' (p.50). But if, at the same time, the construction of the original position and the derivation of the principles must fit some prior moral judgements (even if only those of the reader and the author), the only way to interpret the nature of those judgments is to take them as a product of moral intuitions. But in that case the social contract has no independent justificatory force but is merely an expository or didactic device for explaining principles already held.

Consider a simple example. Two young men meet to play tennis. It is a sunny morning and one of them has to play with the sun in his eyes, which is an obvious disadvantage. For the sake of this argument let us assume that there do not exist (or the players do not know about) the generally accepted rules to do with alternating ends. These two players are therefore to construct just principles for the purpose of their match. After the first game A, who happened to play looking into the sun, proposes a change. Their dialogue:

A: All right, now it's time to change ends.

B: Why? We have not agreed to any rule about changing ends so let us continue playing as we did up to now.

A: It is true that we did not agree expressly but I took it for granted that we would change. After some time we will change again and so on.

B: Why did you take for granted such a rule? *I* did not and therefore I have no duty to do something I haven't agreed to.

A: You have no duty but it is the only possible just solution in a situation when one of us has to play looking into the sun.

B: What is just about it?

A: Well, imagine that before we started playing we sat down to discuss a just solution to this problem (something that, as I see, we should have done, but it is not too late because we can try to imagine our possible discussion). Do you think that, if we reached an agreement acceptable to both of us (in a situation where neither of us knew who would actually play first looking into the sun), such an agreement could be called just?

B: Certainly, under the condition that both of us would have good reasons to agree to it.

A: Fine. Now look: we *knew* in advance that one of us would have to play looking into the sun but we did not know who it would be. We would also agree that there would be nothing just in solving this problem by *faits accomplis* that is, by taking the better end by surprise, force, etc. Right?

B: Right.

A: In this case we would like to design a procedure which would evenly distribute the burden. No one would then be disadvantaged overall. Right?

B: No.

A(surprised): Why not?

B: Because such guidelines for designing a procedure would be based on some *a priori* assumption to which we had not agreed in advance. It would be, in this case, the assumption that some of the burdens of playing should be equally distributed.

A: Yes, but it is of the essence that in sport all participants should face equal obstacles, burdens etc., so that the person who is physically or technically better, wins.

B: No, it is not of the essence. It is only *your* view about the nature of sport.

A: But what is the sense of a sporting contest where the contestants do not have an equal chance of winning?

B: That is another matter, subject independently to agreement. Besides, your views about equal chances are controversial. Imagine that before our discussion we knew that one of us was a better player. Even if we did not realise who, we had good reason to believe that in the ranking list of our club one of us had recently fared better than the other. We do not know who, but we can check it after reaching agreement in principle. In such a case, it is just that the one (whoever it is) who is lower on the ranking list, play at the better end of the court. In this case the one, presumably the least skillful, has an equal chance and only in this way can the equality of opportunity postulated by you, be implemented.

A: That is ridiculous. Such a principle, if generally adopted, would totally distort all sporting competitions. Do you suggest that slower sprinters should run half the distance only, or that stronger boxers should be allowed to use only one hand? All participants to the competition should be bound by the same rules, should face similar challenges and difficulties: otherwise sport would cease to be sport!

B: No, it would cease to sport as you understand it. For me the real nature of sport lies in the sheer pleasure of the game: this aim is best served when actual opportunities are equalised, and that requires taking into account also differences in skill. And this assumption suggests that the rule I propose (and would have proposed if we had discussed it before the match) would be: whoever is more skillful, plays looking into the sun.

I believe that this passage illustrates that one and the same contract situation (and neither of the players questions the conditions of the hypothetical 'original position') can be a basis for two (or more!) completely different sets of principles as a function of different prior evaluative assumptions. In Rawls' case, those are prior moral assumptions that cannot have any other source than intuition. Consider his idea of the lexical priority of liberty: a rule saying that no amount of socio-economic gain (to which his Second Principle of Justice applies) is a good reason for accepting less extensive or less than equal liberty (described by the First Principle). This priority rule is derivable from the original position only if some additional moral evaluations are accepted, which are not inherent in the initial contract situation itself. It does not follow *solely* from a rational choice situation with its 'weakest possible assumptions'. Although Rawls several times (*inter alia*, at pp. 542-43) promises that he will show *why* at a certain stage of socio-economic development after the most urgent material needs are satisfied, 'it becomes and then remains irrational from the standpoint of the original position to acknowledge a lesser liberty for the sake of greater material means and amenities of office' (p.542) he actually does not honour this promise. Nowhere in the passages dealing with the priority of liberty does he explain *why* it would be *irrational* for the original contractors to design a scheme of trade-off between liberty and economic gain. Whenever he asks himself a question about the grounds of the priority of liberty, he answers by asserting that liberty (after basic needs are satisfied) is more important than material gains. Time and again he repeats phrases like this: 'Increasingly it becomes more important to secure the free internal life of the various communities of interest' (p.543) or stresses 'the central place of the primary good of self-respect and the desire of human beings to express their nature in a free social union with others' (p.543) etc. But this is simply the assertion of the moral primacy of the value of freedom over other values, and in particular, an assertion of the principle that no amount of economic gain can compensate any loss of liberty. Those

assertions are no 'justification' of the priority of liberty in non-tautological terms - that is, in terms other than the supreme value of liberty. But Rawls repeatedly suggests that this principle would be adopted by the original contractors as the most reasonable and, therefore, that it can be justified by appeals to values other than those directly postulated by this principle (otherwise the argument would be circular).

The most likely candidate for those other values would be rationality understood as prudence. In a characteristic 'justification' of the two principles of justice, Rawls says that:

> the two principles of justice have a definite advantage. Not only do the parties protect their basic rights but they insure themselves against the worst eventualities. They run no chance of having to acquiesce in a loss of freedom over the course of their life for the sake of a greater good enjoyed by others (p.176).

But again, it is a restatement of a position rather than an argument. The last two phrases justify a view of the 'advantage' of Rawlsian principles only if we assume that parties to the original position will absolutely prefer freedom to material gains in case of conflict between both. But perhaps they are less seduced by the vision of the most extensive possible freedom than by the possibility of some additional socio-economic gains; perhaps they prefer rather to 'run no risk' of sacrificing the most extensive possible economic benefits. In this case it would be reasonable for them to approve a loss of a portion of their freedom 'for the sake of greater good enjoyed by others' if the scheme satisfies the Difference Principle, that is, if greater good is enjoyed by the worst-off. The priority of liberty is therefore being assumed rather than justified; it is one of the tacit evaluative assumptions of his theory that cannot be considered as one of 'the weakest possible assumptions'.

If that is correct, then it is hard to see why the original contractors would find this view necessarily the most reasonable view without making some prior judgments. But, in the Rawlsian

social contract, such prior judgments have no moral force. Rawls stresses that 'in their deliberations the parties [to the original contract] are not required to apply, nor are they bound by, any antecedently given principles of right and justice' and thus that 'there exists no standpoint external to the parties' own perspective from which they are constrained by prior and independent principles in questions of justice'[14]. If such prior moral constraints do not exist, the derivation of principles of justice from the original position (as the most reasonable set of principles derivable from this position) is untenable. This conclusion cannot but be strengthened by Rawls' own account of the original position in terms of 'pure procedural justice' in which there exist no independent criteria of just outcomes but only criteria of just procedures. With regard to a social contract in the original position '[t]his means that whatever principles the parties select from the list of alternative conceptions presented to them are just'[15]. Now, if they are not bound by any prior constraints then there is no reason to believe that they will choose the priority of liberty or any other elements in his conception of justice. Everything of course depends on *who* presents to the contractors the list of alternatives and by what constraints *he* is bound. Now, if the original position is designed in such a way as to exclude the possibility of choice of some principles contrary to prior moral judgments, then the parties to the original contract are nothing more than puppets moved by the invisible hand of the author and are bound by *his* moral constraints. But that only pushes back the matter of pre-contractual, substantive moral intuitions from the position of the 'contractors' to the position of their creator.

To be sure, there is in Rawls a device for establishing a coherence between intuitive moral judgments and principles of justice, as agreed to in the original position. This is a 'reflective equilibrium': a two-way deliberative movement between our considered convictions of justice and the principles derived from the original position. In order to achieve equilibrium, we may either change the circumstances of the original position (so that the principles derived from it match our intuitive judgments) or modify our con-

victions when we are not very confident about them. Therefore, in the case of very strong moral convictions (such as, Rawls says, that religious intolerance or racial discrimination are unjust) we would rather manipulate the conditions of the original position in order to derive principles that are consistent with these judgments. In other cases (such as, Rawls suggests, distribution of wealth and authority), we will look to already established moral principles to find some guidance in these matters. Equilibrium is achieved when our principles and judgments finally coincide.

This view seems to recognise the role of intuition quite expressly. In its 'from judgments to principles' part, the reflective equilibrium appeals directly to our strongly held judgments which cannot be derived from the original position because it is the latter that is scrutinized using those considered judgments:

> [w]e can check an interpretation of the initial situation... by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed (p.20).

So far so good; a problem arises over where the distinction between the judgments about which we are 'confident' (and which therefore serve as a test for the original position and principles derived from it) and those about which 'we have much less assurance' (and which are modifiable to comply with the principles) is to be drawn. The distinction between more and less strongly held principles determines the direction of argument in the reflective equilibrium: in the first case it is 'from judgments to principles', in the second: 'from principles to judgments'. The evident appeal to intuition is only in the first type of reflective-equilibrium procedure; in the second type the appeal to intuition is indirect, *via* the coherence argument (that is, the coherence of judgments about which we are uncertain with the principles that we have already firmly established). But the line between the first and the second type of judgment, cannot be drawn but by an appeal to intuition. Rawls, as a matter of examples, considers the ban on

religious intolerance and racial discrimination as strongly held moral convictions ('fixed points of our considered judgments of justice') while views about the correct distribution of 'wealth and authority' give rise to moral doubts and are held without condifence (pp.19-20, cf. p.206). But we can very well imagine a person being in a quite different situation: confident about the morality of a certain type of political structure and wealth distribution, but hesitating about the justification and the scope of religious tolerance. To suggest therefore a moral system in which a line between those two types of judgment can be drawn is again to rely heavily on intuition which would determine the different levels of certainty of various different judgments.

There is probably nothing particularly revealing in showing that a conception of justice relies on moral intuitions. No philosopher is a magician who can derive substantive moral principles from more or less neutral assumptions. Rawls himself admits that 'any conception of justice will have to rely on intuition to some degree' (p.41). But if this 'degree' turns out to be very high, the justificatory power of the contract argument becomes questionable. Let us sum up: one half of moral reasoning ('reflective equilibrium') appeals directly to intuitions, the other half - indirectly (insofar as the conditions of the original position are tailored to fit moral intuitions). It is hard to see any room for a contract argument as an independent source of moral opinions. Rawls, however, very clearly attributes such an independent justificatory role to the contract: 'certain principles of justice are justified because they would be agreed to in an initial situation of equality' (p.21). But this 'initial situation' itself is constructed so that it matches our convictions; it cannot therefore serve as an independent justificatory device.

Now, 'reflective equilibrium' *does* have an important role in actual moral reasoning as a device for bringing about coherence of moral principles and particular moral judgments. It is also a fair description of our usual reasoning about hard moral cases: we try to work out a general principle on the basis of our strongly held particular judgments (inductive reasoning) and then apply this moral principle to a difficult or controversial issue (deductive

reasoning). In the situation of a clash between a general principle (established on the basis of some other particular convictions) and a particular conviction of justice, we must have freedom either to modify the general principle or to change the particular judgment of justice. Aversion to changing the principle when it clashes with strong moral convictions is a symptom of moral dogmatism; incapacity to change judgments of justice when principles so dictate is a sign of an unprincipled, case-by-case approach to morality. In its extreme version, this latter approach deprives moral discourse of its essential feature: the universality of moral principles, the defence of a moral decision in terms more general than one particular case. Those two deviations should be avoided by a two-way approach to a 'reflective equilibrium', but a decision about which way should be chosen in any particular instance is to be made by the person who makes the moral decision or proposes the moral principle. Ultimately, it depends on the relative intensity of our convictions and principles. If, however, we decide to modify a principle under the influence of a very strong particular judgment of justice, we must remember that its change will lead to a change in some other particular judgment of justice, on the basis of which we have established the very principle that is subject to a change now.

'Reflective equilbrium' is, thus, a useful and adequate description of the introduction of coherence into a moral system through a combination of inductive and deductive reasoning. But, to avoid the objection of circularity, the 'equilibrium' must have its 'external' source. It is not a satisfactory argument if we say, for instance: racial and religious discrimination are immoral therefore all discrimination on irrelevant grounds is immoral (induction); discrimination of women is based on irrelevant grounds therefore it is immoral (deduction). Although that is probably the way we actually often think, a moral theory tries to find some justificatory source for the whole argument. Such an 'external' source of reflective equilibrium in Rawls' theory is provided by a hypothetical social contract but, for reasons suggested above, I

think it cannot be seen as an independent moral source because the conditions of the contract are already moulded by the convictions of justice that are believed to be a part of 'reflective equilibrium'. What is the source of those convictions that influence the conditions of the initial contractual situation? Answering this question Rawls, as I tried to show above, must appeal either directly or indirectly to intuition. Hence social contract is superfluous as a part of moral justification: it can be viewed as a useful didactic tool, but as a part of 'reflective equilibrium' it creates an illusion that it is an 'objective', non-intuitive factor in justification. In a recently published article Rawls declares that:

> [t]he aim of political philosophy ... 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, ... to propose to it certain conceptions and principles congenial to its most essential convictions and historical traditions. [16]

It is clear then that for Rawls common sense precepts, essential convictions and historical traditions are *prior* to, not derived from, conditions for a social contract. They shape the structure of the original position, not the other way round.

The upshot of these considerations is that, ultimately, we appeal to our moral intuitions when proposing principles of justice and that the job done by a hypothetical social contract is derivative from prior moral intuitions. Now, this view may seem very unsatisfactory to those who seek more solid and 'objective' grounds for principles of social justice than something as enigmatic and unintelligible as moral intuitions. Secondly, this view may also be objectionable to those who are offended by the apparent absolutism of 'intuitionism' conceived as a theory claiming the discovery of self-evident moral truths by means of 'direct insight'. Paradoxically, then, direct appeal to moral intuition is susceptible to quite opposite objections: that it is too weak and too strong,

that it does not help solve moral controversies and that it attempts to impose arbitrarily 'true' answers to difficult moral questions.

To take up the second point first, I should like to suggest that the ultimate appeal to moral intuitions in defending moral principles and judgments does not necessarily entail endorsement of a meta-ethical theory of 'intuitionism' with all its traps and overstatements. In particular it does not necessarily entail two views shared by meta-ethical intuitionists, that, firstly, truth and falsity can be attributed to moral judgments and hence that morality is a matter of knowing (although in a different sense than knowledge about natural facts), and secondly, that moral truths are self-evident, necessary and indubitable because they are propositions synthetic *a priori*.[17] Without entering into the debate on other merits or demerits of such radical intuitionism, it seems to me that the view endorsing these two points has a practical setback: it clouds rather than illuminates the seriousness of moral disagreements about judgments and principles of justice. If the persistence and the inevitability of such disagreements is taken to be one of the most important facts about morality, and in particular about judgments of social justice, then meta-ethical intuitionism is hardly helpful in explaining the nature and the sources of such disagreements. Intuitionism, endorsing the two propositions cited above, cannot avoid the conclusion that in cases of moral controversy some people know the truth, others are mistaken. In consequence the intuitionist must end up with a statement that:

> the fact that people disagree in moral matters, even concerning basic, ultimate moral issues, is evidence that they cannot all be right, not that the judgments involved are incapable of truth or falsity.[18]

This view, although theoretically coherent, is of no help when actual moral disagreements arise. If, on the one hand, people's moral judgments are thought to be a matter of truth and falsity and, on the other hand, those judgments obviously differ among themselves,

how are we to know whose intuition is capable of discerning the moral truth and whose intuition is deficient? Intuitionism postulates that there is a moral truth and that it is discernible by intuition but it fails to help us to select the moral truth from among the variety of moral judgments. How can we distinguish a genuine moral intuition from a false one?

To be sure, an intuitionist *does not have to* answer this question. It is theoretically coherent to hold the view that moral propositions are about the truth and at the same time that this truth is unverifiable in cases where disagreement arises. The very fact of the existence of moral disagreements is therefore too weak an argument against intuitionist meta-ethics, contrary to what some of its critics claim[19]. But it is a sufficient argument against the usefulness of this theory. After all, the most obvious question that arises about the intuitionist claim that moral truth is self-evident is: self evident - for whom? Actual moral disagreements show that this 'self-evidence' must be represented by some special inner faculties of persons whose intuitions are better than those of others - but this is a path that very few contemporary intuitionists would like to follow.

The other type of objection to intuitionism is put forward by Rawls. His main argument is that intuitionism is of no help in weighing moral principles against one another in case of conflicting values, because it contains no priority rules and 'we are simply to strike a balance by intuition, by what seems to us most nearly clear' (p.34). This 'no-priority-rules' argument is by far the most important objection that Rawls formulates in his polemics against intuitionism: we are told that '[t]he intuitionist believes ... that the complexity of the moral facts defies our efforts to give a full account of our judgments and necessitates a plurality of competing principles' (p.39). However, what Rawls is arguing against is but one possible version of intuitionism, rather unrepresentative and clearly implausible. The view that he is refuting is that moral intuitions (which are sources of judgments) cannot suggest any priority rules in case of competing values or principles: no principle

is ultimate, there is a plurality of second-to-ultimate principles and conflicts among them are practically solved only on a case-by-case basis. According to Rawls:

> Intuitionism holds that in our judgments of social justice we must eventually reach a plurality of first principles in regard to which we can only say that it seems to us more correct to balance them this way rather than that (p.39).

That is, however, to attack a man of straw. This version of intuitionism, although conceivable, is an incomplete conception of justice. It states that there are several 'first principles' (in fact, they are only 'second principles') but by failing to prescribe priority rules it is useless for guiding our actions in hard cases. There is no reason to think that a meta-ethical theory of intuitionism must be linked to an incomplete conception of justice: intuitionism may, just as any other theory, postulate priority rules. If we can intuit several 'first principles', why can we not intuit *the* first principle? There is no reason for moral intuition to stop at an intermediate level in the hierarchy of moral principles. Actually, Rawls himself observes: 'Perhaps it would be better if we were to speak of intuitionism in this broad sense as pluralism' (p.35). This is a disarming statement: of course it would be better! But in this case Rawls' objections would hold against *any* form of moral pluralism that might just as well be a social contract theory as an intuitionist one, and then it would become obvious that his arguments are not relevant as a criticism of intuition as a moral source. Hare rightly observes that:

> [t]here can also be another, non-pluralistic kind of intuitionist - one who intuits the validity of a single method, and erects his entire structure of moral thought on this.[20]

Rawls' arguments about the weakness of intuition are therefore misdirected: they also probably stem partly from a confusion of the meta-ethical theory of intuitionism with the normative intuitionist

ethics that is actually refuted (in one of its versions) by Rawls.[21] Intuitionism's weakness lies elsewhere: not in the fact that it does not guide our actions by clear priority rules but that the scope of interpersonal arguments about them is so limited. People's values, including principles of justice, are not, as emotivists maintain, solely expressions of their emotions or recommendations of actions. They are not totally 'arbitrary' or 'irrational'. Principles of justice are often the result of rational considerations about the possible consequences of various rules, structures and actions; the consequences for human beings in terms of their life, dignity, prosperity, liberty etc. People value certain principles more than others because they *know* (or think that they know) possible consequences of putting them into practice - that is a 'rational' part of arguments about justice and that is a legitimate field for moral dialogue. Moral judgments are not derived from facts but they correspond in a certain way to facts. People make their judgments not always arbitrarily and not always as emotional responses to challenges but also as considered convictions based upon expected consequences resulting from acting upon them. But the ultimate evaluation of those consequences hangs upon subjective principles: standards of right and wrong underivable from empirical facts or from even higher standards. Those ultimate value judgments cannot be argued about in terms of empirical facts because it is the latter that are assessed by the former. Reflection upon the facts and the facts *alone* does not entitle us to make value judgments about them. In this sense the appeal to intuition is subjectivist without being emotivist and without - on the other hand - necessitating any claim to discovery of self-evident 'moral truth'. It is the view that ultimately any moral disagreement is reducible to a statement of opposite values which are neither arbitrary expressions of emotion nor the opposition of truth and falsity. Appeal to intuition is an admission of the limits, not the impossibility, of reason in moral matters.

The alternative: either our value judgments express 'moral truth' and therefore in case of disagreements at least one view is mistaken or our judgments are totally arbitrary and irrational, does not exhaust the whole list of possible approaches to morality. The view that I would like to suggest is that moral judgments are neither arbitrary nor verifiable in interpersonal discourse. The fact that people cannot prove the truth of their principles of justice and that those principles cannot be attributed 'truth' from any human point of view, does not necessarily mean that those people hold their respective views without any rational justification. Those judgments cannot be 'proved' but they can be justified; they can be justified, but they cannot be agreed upon by all. Not because some of them are less strongly justified, or because some people are less rational or more partial etc.: the lack of moral consensus is not a contingent, but an inherent feature of human morality. Human disagreements about considered principles of justice and the impossibility of reaching any consensus in interpersonal discourse, express the very nature of morality.

NOTES

* I wish to thank Mr. Roger Wilkins for his important comments on a draft of this paper and for his help in correcting my English.

1. The subsequent parenthesised references in the text are to John Rawls, *A Theory of Justice*, (Oxford: Clarendon Press, 1972)

2. Ronald Dworkin, 'The Original Position' in Norman Daniels, ed., *Reading Rawls* (Oxford: Basil Blackwell, 1975), p.18

3. Examples of such confusion and, in consequence, unjustified criticism of Rawls: Allen Bloom, 'Justice: John Rawls Vs. the Tradition of Political Philosophy', *The American Political Science Review* 69 (1975), pp. 648-62; David Lewis Schaefer, *Justice or Tyranny? A Critique of John Rawls' "Theory of Justice"* (Port Washington: Kennikat Press, 1979), pp. 39-41 Scheafer's book is, by the way, a genuine anthology of demagogical and unfair criticisms.

4. John Locke: *Two Treatises of Government*, Cambridge: Cambridge University Press, (1970) II, 104, p. 354.

5. *Ibid.*, II, 114-121 (pp. 362-67).

6. Jean-Jacques Rousseau, *The Social Contract and Discourses*, translated by G.D.H. Cole (London: J.M. Dent & Sons, 1973), p. 165, emphasis added.

7. Thomas Hobbes, *Leviathan* (London: J.M. Dent & Sons, 1928), p. 90

8. James Buchanan, *The Limits of Liberty* (Chicago: The University of Chicago Press, 1975), p. 26, emphasis added.

9. *Ibid.*, p. 25

10. *Ibid.*, pp. 23, 28

11. *Ibid.*, p. 167

12. *Ibid.*, p. 73

13. Cf. Norman Furniss, 'The Political Implications of the Public Choice - Property Rights School', *The American Political Science Review* 72 (1978), p. 407

14. John Rawls, 'Kantian Constructivism in Moral Theory', *The Journal of Philosophy* 77 (1980), pp. 523-24.

15. *Ibid.*, p. 523

16. *Ibid.*, p. 518

17. H.J. McCloskey, *Meta-Ethics and Normative Ethics* (The Hague: Martinus Nijhoff, 1969), ch. 4 and 5

18. *Ibid.*, p. 143

19. Cf. Richard Brandt, *Ethical Theory* (Englewood Cliffs: Prentice Hall, 1959), pp. 192-96; John Hospers, *Human Conduct* (New York: Brace & World, 1961), pp. 539-40

20. R.M. Hare, 'Rawls' Theory of Justice' in N. Daniels, *op. cit.*, p. 84.

21. On the distinction between normative and meta-ethical intuitionism, cf. Jonathon Harrison, *Our Knowledge of Right and Wrong* (London: Allen & Unwin, 1971), pp. 74-6.