JUSTICE, SOCIAL CONTRACT, AND UNIVERSAL PRESCRIPTIVISM

BY M. H. LESSNOFF

Among the many distinguished critics of Rawls’ theory of justice, one of the severest is Professor R. M. Hare.¹ Hare’s critique of Rawls seems to me often captious and sometimes obscure, but the relation between the views of the two men is, I believe, worth examination, not least because it turns out to be an unexpectedly ironical one. For Hare, in this connection, maintains the following positions: (1) Rawls’ premise—that principles of justice are to be found in the agreement of rational self-interested persons ignorant of their individual peculiarities—is precisely equivalent to that of his own moral theory of universal prescriptivism; (2) each of these premises, properly handled, yields utilitarian conclusions about the content of morality, including justice—and not the “maximin” rule that bulks so large in Rawls’ theory.² I shall try to show that in fact Hare’s premise (universal prescriptivism), at least as he interprets it, yields Rawls’ conclusion (maximin), whereas Rawls’ premise (his “original position”) yields, or at least does not exclude, Hare’s conclusion (utilitarianism). It follows, since utilitarianism and maximin are incompatible, that the respective premises of Rawls and Hare are not equivalent. Whether one should accept either premise, or the conclusions derivable from it, is another question.

1. RAWLS AND SOCIAL CONTRACT THEORY

Rawls’ original intuition, the basis of his theory, seems to have been that the relationships and constraints inherent in the idea of justice are well captured by the traditional concept of the social contract. In course of time, however, Rawls found it necessary to modify that concept.³ Let us try to see why, and to assess the result.

The notion of contract is closely connected with that of exchange. A contract between two parties, says the O.E.D., is “a mutual agreement that something shall be done or forborne by one or both”. Most frequently, it is by both: that is, each party undertakes to act or forbear, given that the undertaking of the other party is also fulfilled. Such a contract is an exchange of conditional undertakings—usually, of undertakings to perform services. That is, the actions or forbearances undertaken by each party are to the advantage of the other party, and likely, in themselves, to involve some cost or sacrifice to the undertaking party. For each party, the carrying out of the contract involves a balance of cost and benefit, a balance which, if the contract is freely made, must be presumed to be favourable.

² Op. cit., pp. 152-5, 246-51. See also “Rules of War and Moral Reasoning”, Philo-
Both parties gain from the exchange of services involved in the carrying out of a free contract; indeed, that the parties do thus gain can be taken as the criterion of the contract's being a free one. In economic theory, the gains that accrue from free contractual exchanges of services are called "gains from trade".

In the philosophical concept of the social contract, this idea is extended from two parties to many. In the theories of Hobbes and Locke, for example, each individual party (to the "covenant" or "compact") undertakes to give up a right (the natural right to every thing in the case of Hobbes, the executive power of the law of nature in Locke's), or, more exactly, to transfer it to a central governmental authority. This renunciation is a sacrifice by the individual, more than outweighed by the benefits he receives from the similar renunciations of his fellow-contractors.

At first glance, the free contract model may indeed seem promising as a paradigm of justice. If what is agreed is freely agreed, and benefits all parties, is it not ipso facto just? But that would be an over-hasty conclusion. As is well known to economists, the gains from trade between two parties can be distributed between them in an indefinitely large number of ways, depending on their relative bargaining power. Marx argued in his analysis of capitalist profit that free contract is thus perfectly consistent with exploitation. Under conditions of "perfect competition", there are so many buyers and sellers of a given service that bargaining power is by definition non-existent, and a single price reigns in the market. But even this does not mean that the gains from trade are justly distributed; rather, their distribution reflects the market value of the resources individuals happen to have, and are willing to market.

These conclusions apply equally to the philosophical concept of the social contract, as we can readily see by considering the predicament of a Lockean individual in a Hobbesian world. The Lockean individual accepts and would abide by the law of nature: he accepts that "no one ought to harm another in his life, health, liberty or possessions". But living in a Hobbesian state of nature, he is surrounded by individuals who have no respect for such rights, and his life is likely to be nasty, brutish, and short. Clearly he will gain by the Hobbesian covenant which will (we may agree with Hobbes) protect his life, if not his liberty or property. But this is not the social contract he would prefer to make—his preferred contract would be something like Locke's version, in which government acquires only a limited fiduciary power to protect the natural rights of its subjects. In a Hobbesian world, however, this option is not open to him, for the Hobbesians who surround him will agree to a commonwealth headed by a Hobbesian absolute sovereign, and to that only. The Lockean has the choice of acceding to this covenant, or of remaining in a state of war with the commonwealth—and clearly he must accede. He benefits thereby, and yet is the victim of injustice. For he would be willing to respect the natural rights of others, if they would respect his. But they are not—or at least, not willing to limit the government's authority in order to safeguard them—and so his natural rights are not safeguarded, and may not be respected.

Social contract as such, then, does not suffice to ensure justice. Rawls' strategy, however, is to modify the concept for his own purposes. One modification is obviously necessary, in the content of the contract—the

---

4 See Leviathan, chs. 14 and 17; Second Treatise of Government, sections 7 and 87.
5 Locke, op. cit., section 6.
contractors are to agree on principles to govern the distribution among them of primary social goods. Besides this, they are to be not only equal in bargaining power, but unable to take advantage of any special command of valuable resources, because they are unaware that they have any. In this sense the contractors are all identical atoms, though they know that in the real world (when the "veil of ignorance" is removed) different people are differently endowed. How successful are these moves in producing a theory of social contract suited to the generation of principles of justice?

This question is really two questions: (1) is Rawls' "hypothetical construction" still recognizably a theory of social contract? (2) is it suited to the generation of principles of justice? On the first question, it is necessary to compare Rawls with Hobbes and Locke (the "classics"). One element of the classical theories, one which is inherent in any contract of exchange, seems to be lacking in Rawls' theory—namely, some conception of the "non-contract situation". In the classical theories, this is the so-called state of nature—it does not matter whether this is conceived as historically prior to the contract, or merely as a hypothetical state which would obtain if the social contract were not adhered to. Either way it is a description of the alternative available to the contractors, which is less desirable to them than general adherence to the social contract. It is because of this alternative that the social contract can be seen, for each contractor, as a sacrifice outweighed by greater benefits.

Rawls does not seem to describe such a non-contract alternative. Certainly he does not dwell on it to anything like the extent of Hobbes and Locke on the state of nature. It is true that he does remark in passing that "general egoism" is "what the parties would be stuck with if they were unable to reach an understanding". By general egoism, Rawls means a conceivable (though not acceptable) principle for the ordering of social interaction, according to which "everyone is permitted to advance his interests as he pleases". At first glance, this seems very similar to Hobbes' natural right to all things, and so it may seem that Rawls is here postulating, as his non-contract alternative, a Hobbesian war of all against all. But this, I think, is not so. Rawls' notion of general egoism is only a possible rule (or the absence of rules) to govern conduct: it is not a conception of human nature. There is no suggestion in Rawls' theory that general egoism is even an approximately accurate description of human behaviour or social life in the absence of a contract, on the basis of a conception of human nature. The premises of the Rawlsian contract are, in a sense, partly Hobbesian: that is, his contractors, like Hobbes', are thoroughly self-interested individuals. But in Rawls' case this characteristic is postulated in order to reflect the "circumstances of justice", whereas Hobbes takes it to be a sufficiently accurate description of the nature of men.

The absence of any clear conception of a non-contract alternative suggests that Rawls' theory is not genuinely a social contract theory. This is not to say that Rawls' theory does not embody a contract, as defined by the O.E.D.; only that, unlike the classical social contract, it is not a contract of exchange. But now I want to consider the second question, whether the quasi-Hobbesian premise of Rawls' theory suffices to generate principles of justice. A striking feature of Rawls' "social contract" is that in it quasi-

*A Theory of Justice, section 24.


Hobbesian premises are used to generate quasi-Lockean conclusions. To be more explicit: in Rawls’ theory thoroughly self-interested individuals, recognizing no obligations to others, come to accept the constraints of a morality—a morality of individual rights, moreover, which they undertake to respect and which the basic structure of society is to protect. In Locke’s version of the social contract, on the other hand, while the parties do similarly respect a morality of individual rights and establish social machinery to protect them, that morality is not generated by the contract—rather it is antecedently accepted by the parties as the “law of nature”. In the Hobbian covenant, the self-interested, amoral contractors do not institute any protected individual rights, but rather an absolute sovereign. Hobbes’ conclusion, however, may seem unconvincing. Since his contractors wish above all to avoid violent death, would it not be rational for them to institute a right to life similar to Locke’s—one that is to be protected, not only against other private individuals, but against the sovereign himself? Why does Hobbes not limit the sovereign’s authority in this way? There seems to be only one possible explanation, namely, that he thinks that the lives of subjects are actually better protected by absolute sovereignty than by an entrenched Lockean right to life. This is not to deny that absolute sovereignty carries with it some risk to subjects’ lives; however, Hobbes presumably held that any diminution of the sovereign’s authority carries a greater risk because it increases the likelihood of disorder. In other words, the Hobbsian conclusion apparently rests on a balancing of probabilities in a situation of uncertainty, and in this respect is similar to Rawls’ argument (see below). I suggest that such an approach is, contrary to Rawls, intrinsically unsuited to the generation of just outcomes, and to the entrenchment of individual rights that justice presumably requires. For let us suppose that Hobbes’ apparent estimate of the relevant probabilities is correct: a subject’s life is likely to be best protected by absolute sovereignty. It is not for this reason any less of an injustice (whatever Hobbes may say), if in a specific instance a sovereign for some reason decides to have an innocent subject killed.

This leads to a major criticism of the Rawlsian argument. Rawls’ contractors are, as explained above, identical atoms unaware, as contractors, of their differing endowments as social men. By this move Rawls converts his “social contract” into a situation of choice under uncertainty—choice, that is, between hypothetical societies structured in accordance with alternative possible principles of justice, under uncertainty as to what position the choosing individual will occupy in any of these social structures. Since Rawls has rendered all his contractors identical, the same choice situation is posed to all of them, so that, as Rawls notes, we need consider the rational preference of one individual only.10 This preference must involve a balancing of probabilities, and is for that reason unsuited to the generation of principles of justice and of the entrenched rights stipulated by Rawls’ own preferred principles. In general, Rawls seeks to avoid this conclusion, but in one place at least his own argument unwittingly demonstrates the problem. In §38 Rawls argues that, given that a certain sort of social structure is preferred by his contractors, they will also prefer that the rules defining that social structure be enforced by law, so that the structure in question is


10A Theory of Justice, p. 139.
stabilized. He then raises the question whether such laws may include laws of strict liability, that is, laws prescribing that penalties be inflicted on individuals on account of facts which these individuals did nothing to bring about, and did not negligently fail to prevent. His answer is that, in some cases, such laws are justified.\footnote{Op. cit., pp. 242-3.} For example, in a situation of potential civil disorder, it would be right to punish persons in whose homes weapons are found, even if these persons neither knew of nor intended the presence of such weapons; it would be right to do so if the penalties were not too severe, and considerably improved the chances of preserving public order. As Rawls puts it,

the formation of paramilitary groups, which [such a] statute may forestall, is a much greater danger to the freedom of the average citizen than being held strictly liable for the possession of weapons. Citizens may affirm the law as the lesser of two evils, resigning themselves to the fact that while they may be held guilty for things they have not done, the risks to their liberty on any other course would be worse.

Such hypothetical reasoning on the part of citizens seems sound enough, but does not show that it is ever just (as distinct from justified) to punish people for things they have not done. Someone actually punished in the way Rawls describes is certainly the victim of injustice, though quite possibly his treatment is justified as the only way to prevent a greater evil, and even, as Rawls suggests, worse injustice.

Now, the bearing of this argument on Rawls' theory in general is the following. Rawls does not deny that in the case of strict liability injustice is done, that a person's just rights are infringed if he is punished for something he has not done. However, rational acceptability from the point of view of the self-interested "representative citizen", which Rawls gives as the justification of strict liability, is precisely what is supposed to ensure the justness of his principles of justice. If, however, it fails to do so in the former case, it presumably fails in the latter case also. In the strict liability case, Rawls' representative citizen has to balance certain risks (probabilities of harm) and thus cannot exclude the risk of harm—in fact, of suffering injustice. The Rawlsian contractor, asked to choose between alternative social systems in ignorance of what would be his position in them, likewise must weigh his chances of doing well or ill in these systems. As many of Rawls' critics have pointed out,\footnote{For example, Hare in "Rawls' Theory of Justice", op. cit., pp. 246-51; B. M. Barry, The Liberal Theory of Justice (Oxford, 1973), ch. 9.} he does not succeed in showing that it would not be rational for his representative individual to prefer a society that maximizes average utility (since the average is what an individual unable to distinguish himself from anyone else should expect to receive) to one that (as Rawls would have it) maximizes the minimum quota of primary social goods. Rawls' premises, in other words, can lead to a fairly standard kind of utilitarianism. But this does not show that a society that maximizes average utility is just; it is compatible with gross inequalities, severe and unmerited deprivation, even, as Rawls himself has cogently argued, with slavery. A Rawlsian contractor, though not irrational in opting for such a society, would risk suffering severe injustice when actually in it, since it might be a very unjust society.

Against this argument a partial defence of Rawls might be attempted,
along the following lines. While it is not necessarily irrational to take risks, and so Rawls fails to establish that his rational self-interested contractors must prefer the maximin distribution of primary social goods (which is the no-risk solution of the choice problem), nevertheless there must be some limit to the risks which it is rational to run. Surely it would not be rational to risk one's life or basic liberty for the sake of the chance of a higher quota of other social primary goods? That is, would it not be rational to prefer a society which guarantees the Lockean rights of life and (basic) liberty, as well as a right to a level of income sufficient for subsistence (and even perhaps for a modest degree of comfort), to one which does not, whatever the level of average utility in the two societies? And are not a right to life, basic liberty and subsistence (and even perhaps a modest degree of comfort) indeed necessary to a just society? So—the defence continues—while Rawls' premises do not yield his conclusions, are they not nevertheless well adapted to deriving principles of justice?

The defence seems persuasive, but is not quite successful. The postulated choice rests on assumptions that are not so much rational as unheroic. It is not necessarily the case (pace Hobbes) that a rational self-interested person must refuse to risk death for the sake of the chance of other goods—it depends how much he values the other goods, compared with bare survival. It was not irrational for the ancient hero to seek battle for the sake of glory, or for Cortes to seek wealth and power in America at the risk of his life, any more than it is irrational to climb a dangerous mountain in full awareness of the perils of doing so, or indeed to set out on a long ear journey knowing of the possibility of a fatal accident.

I shall conclude this section by pointing to one further consequence of the unsuitability of Rawls' premises for deriving principles of justice. Rawls argues that inequalities in the distribution of income and wealth are just, if they are "to everyone's advantage" (the Difference Principle). The idea here is partly the familiar one of incentives: that is, to offer extra rewards for extra work, for the productive use of special abilities, since saving, risk-taking and the assumption of responsibility may increase the flow of these activities and hence the income and wealth of everyone. Such extra rewards are just, Rawls holds, if they do not exceed the level that maximizes the income and wealth of those who do not receive them. Now, it seems clear that it would be rational for Rawlsian contractors to prefer at least this much inequality (if they were risk-takers they might prefer even more, for the sake of higher average utility), but this does not show that even this much inequality is just. For it may be (I do not argue that it is so, only that Rawls' argument fails to show that it is not) that those endowed with useful skills and abilities ought to make them available for production without extra rewards (along the lines of the Marxist principle "from each according to his ability, to each according to his needs"). If so, and if they never-


14From the point of view of the present argument, Rawls' "basic liberties" are not all sufficiently basic. By "basic liberty" I refer here to what Rawls calls freedom of the person, and freedom from arbitrary arrest and seizure (cf. A Theory of Justice, p. 61).

15It might be objected that Rawlsian contractors, not knowing whether they will turn out to be risk-takers, but knowing that few people are, will be wise to assume they are not. But this, I think, only pushes the argument back a stage.

16A Theory of Justice, p. 78.
theless will not place their useful abilities at society's disposal without extra rewards, and if, further, we decline, in accordance with the Rawlsian priority of liberty, to force them to do so, then their extra rewards under Rawls' Difference Principle are unjust. There ought, according to this hypothetical argument, to be equality of economic reward; true, equality would harm everyone, but only because the well-endowed refuse to use their endowments as they should. If they did use their endowments as they should, the social product would be no less, but would be equally divided, and the poorly-endowed would be better off. We may agree with Rawls that, given human nature, it is better to tolerate inequality in accordance with the Difference Principle than to try to enforce equality with (or without) forced labour, but neither our acceptance of this proposition, nor its acceptability to Rawlsian rational self-interested contractors, affects the possibility that we may be tolerating an injustice for the sake of avoiding greater evils. If it is true that the well-endowed ought to contribute their talents without special rewards, then the position of the poorly-endowed in a Rawlsian society is analogous to that of a Lockean in a Hobbesian world. Both the Lockean and the poorly-endowed are justly entitled to more than they get, and they do not get what they are justly entitled to because of the propensities of their fellow-citizens.

2. Universal Prescriptivism and the Principles of Justice

I have now shown, I hope, that Rawls' contractarian premises do not yield his principles of justice, do not exclude utilitarianism, and are not suited to the derivation of conclusions about justice. I have still to show that Hare's universal prescriptivism leads not, as he asserts, to a utilitarian conception of justice, but rather to a Rawlsian maximin conception.

At first sight, it is hard to see how universal prescriptivism could lead to any particular principles of justice. The requirement that moral judgements be universalizable, as Hare explains, is the requirement that they accord with some universal principle, not any particular one. It is, Hare says, a logical and not a moral requirement. Rawls too requires the principles that are to be considered as candidates for principles of justice to be universal, but correctly notes that this test is passed, not only by his own preferred principles, but equally by the various brands of utilitarianism and by "general egoism". And it does not help to add a requirement of prescriptivity to that of universality—all these principles are equally prescriptive.

The clue to the mystery seems to lie in a small modification on Hare's part to his earlier claim that universal prescriptivism and utilitarianism are "practically equivalent". In "Rawls' Theory of Justice", he describes his theory no longer merely as universal prescriptivism, but as "a combination of the requirement that moral judgements be universalizable and the requirement to prescribe for hypothetical reversed-role situations as if they were actual (I am not sure whether the second is an independent condition or not)". Whether or not the second requirement is an independent condition, it may enable us to solve our problem. To see its significance, we need to consult Hare's earlier Freedom and Reason. Here Hare imagines

18 A Theory of Justice, section 23.
19 Op. cit., p. 93. See also fn. 2 above.
a case in which A owes B money, and B considers whether he ought to have A imprisoned for debt. Hare argues, roughly, that B may sincerely say that he believes he ought to do so only if he can sincerely say that, were the roles of A and B reversed, A ought to have him imprisoned. That is, B is required to prescribe for the hypothetical reversed-role situation as if it were actual, before he is allowed to prescribe for his actual situation.

But how does this help with our problem? So far, no substantial restriction seems to have been placed on what someone may prescribe. However, Hare’s interpretation of his role-reversal condition is that, in the case of A and B, B is forced “to give weight to A’s inclinations and interests as if they were his own”. In principle, B might give zero weight to each person’s interests, but here Hare relies on a psychological generalization, namely that most people wish to further their own interests as much as possible. They can do so on Hare’s conditions, however, only to the extent that they equally further those of others. Where the interests of more than two people are involved, “we find ourselves bound to give equal weight to the desires of all parties (the foundation of distributive justice); and this in turn leads to such views as that we should seek to maximize satisfactions”. For in the multi-person case, the analogue of role-switching in the two-person case is to put oneself successively in the place of all the people affected, giving equal weight to the interests of all.

I shall not comment on the propriety of deriving moral principles from psychological propensities together with a requirement of logical consistency. Rather, I shall now suggest that Hare’s moral theory leads, not to a utilitarian theory of justice, but to a Rawlsian maximin one. For this purpose, let us imagine that a person is asked to choose between alternative principles of justice, subject not to Rawls’ conditions but to Hare’s. How, first of all, would he choose between the suggestion that his society conform to the utilitarian principle that satisfactions be maximized, and the proposal that they be equalized at the highest possible level? If this is a genuine choice, the utilitarian principle must involve some inequality (perhaps, in reality, great inequality)—that is, a state in which some are better off than others. Can a Harean individual prescribe such a principle? Let us apply Hare’s reversed-role test, taking our cue from the case of A and B above. Here the two roles to which we have to attend can be defined, in general terms, as those of better off and worse off, distinguished by whatever universalist criterion is appropriate (say, relative ability). By the reversed-role test, no one is allowed to prescribe a principle that places him in a better position than his fellows, unless he accepts that, if they, rather than he, fell into the relevant category, he should be worse off than they. According to Hare’s psychological generalization, he cannot in practice do this. Nor, of course, can he prescribe a principle that would actually place him among the worse off. Thus, if utilitarianism, as is likely, involves inequalities of this kind, it is not acceptable.

Indeed, it looks as if no inequalities at all are acceptable given Hare’s premises, and this, curiously enough, is a conclusion which at one stage he found attractive.\footnote{Freedom and Reason, p. 121.} However, there is, I believe, one kind of inequality which would be acceptable on Hare’s terms, namely (to use a Rawlsian phrase) an inequality that is to everyone’s advantage. For the Harean universal prescriber, like the Rawlsian contractor, is assumed to be self-interested: in choosing between perfect equality, and inequality which is to everyone’s
advantage, he must prefer the latter—that is, he prefers the Rawlsian maximin principle. (As self-interested, he will want to be better off relative to his fellows than the maximin makes him; but the reversed-role requirement prevents him from prescribing such a principle. Maximin is the best he can get.)

Why is it that Hare’s premises lead to Rawls’ conclusions, whereas, I argued earlier, Rawls’ own premises fail to do so? The reason seems to lie in Rawls’ veil of ignorance, which turns the problem into one of choice under uncertainty. Or rather, it lies in the fact that there is no veil of ignorance in Hare’s theory; instead his universal prescriber is supposed to know where he stands both in actuality and in the reversed-role test. For him, unlike the Rawlsian contractor, there is thus no question of balancing risks or estimating chances, and hence no question of accepting an inegalitarian utilitarianism as a gamble. However, the fact that Rawls prefers the maximin principle, and that it follows from Hare’s premises, does not mean that it is the correct principle of justice. In the case of Hare’s theory it follows directly from his requirement that we must “give equal weight to the desires [interests] of all parties”. This premise is also explicitly embraced by Rawls: his “initial position” is an attempt to build it into the foundation of his theory of justice. It looks, though, as if he would have done better to derive his conclusions directly from that premise rather than through his particular contractarian scenario. At all events, his principles of justice are nothing but a spelling out, in relation to the primary social goods, of Hare’s egalitarian premise.

To start from that premise, however, is to take a great deal for granted—far more than a theory of justice should. Should we in fact give equal weight to the desires or interests of all parties? Whether they are good or bad, whether their contribution to the general good is great or small? These surely are the major questions to which a theory of justice must address itself. Rawls’ theory, no less than utilitarianism and Hare’s universal prescriptivism, fails to do so.

University of Glasgow

---

22Similar reasoning can be applied to the more complex principles of Rawls’ Special Conception of Justice. I discussed above the choice, for a Rawlsian contractor, between (on the one hand) the Difference Principle for distribution of income, plus liberty of the person, and (on the other hand) equal income, at a higher level than the Rawlsian minimum, made possible by forced labour for the well-endowed. A Harean universal prescriber who is one of the poorly-endowed will probably (if we grant Rawls’ priority of liberty, certainly) prescribe the former (as Rawls does) since he is unlikely to be able to accept that, if he were among the well-endowed, he should be subjected to forced labour.

23Freedom and Reason, p. 123.

24A Theory of Justice, pp. 12, 14, 19, 21, 118.