Philosophic Exchange

Volume 10
Number 1 *SUMMER 1979 Volume 10*

Article 6

1979

Bargaining Our Way Into Morality: A Do-It-Yourself

David Gauthier
University of Toronto

DAVID GAUTHIER
Professor of Philosophy
University of Toronto

Bargaining Our Way Into Morality: A Do-It-Yourself Primer

by

David Gauthier

1. "The theory of justice," according to John Rawls, "is a part, perhaps the most significant part, of the theory of rational choice."[1] Let us reflect on the significance of this claim.

Choice is the endeavour to realize one among several alternative possible states of affairs. The rationality which may be exhibited in choice is conceived in maximizing terms. A numerical measure is applied to the alternative possibilities, and choice among them is rational if and only if one endeavours to realize that possibility which has been assigned the greatest number. This measure is associated with preference; the alternative possible state of affairs are ordered preferentially, and the numerical measure, which is termed utility, is so established that greater utility indicates greater preference. The complications of this procedure need not concern us here.[2] What is important is that rational choice is conceived as preference-based choice, so that the rationally chosen state of affairs is the most preferred among the alternative possibilities.

John Rawls' claim, therefore, is that the theory of justice is the most significant part of the theory of preference-based choice. But this claim must seem quite implausible. Justice is a moral virtue — indeed, some would claim that justice is the central moral virtue.[3] The theory of justice must be a part, and perhaps the most significant part, of the theory of morality. How can morality be part of preference-based choice?

The point of morality is surely to override preference. Were we to suppose that one should always endeavour to realize his or her most preferred state of affairs, then what need would we have for moral concepts? Why use the language of morality, of duties and obligations, of rights and responsibilities, when one might appeal directly to each person's greatest interest?

You offer me a choice among pieces of cake. I, greedily but perfectly rationally, basing my choice strictly on my preferences, select the largest piece. "That isn't fair," someone complains. "Of course not," I reply. My concern was not to be fair. My concern was to get the largest piece of cake — and I did. Surely here the appeal to fairness, to a consideration related to justice, is intended to override, or at least to constrain, preference-based choice. If you suppose that I should have chosen with fairness in mind, then you believe that I should not have acted simply to gratify my greed, even though my preference was for the largest piece of cake. You believe that I should have considered, not only my own desires, but also the desires of others.

Do examples such as this show that Rawls is wrong to treat the theory of justice as part of the theory of rational choice? Not at all. I shall argue that his claim is sound. Not that I agree with Rawls' theory of justice — that is quite another matter.[4] But justice provides a fundamental link between morality and preference, a link which, I believe, we are able to formulate in a precise and definitive way.

Indeed, I shall go farther than Rawls. In coming to understand how justice links morality and preference, one also realizes that our framework of moral concepts is seriously outmoded. Morality has been traditionally conceived as embracing

the entire range of justifiable constraints on preference-based choice. But this range will be seen, in the light of my argument, to include at least two distinct, and apparently disparate, parts. One part, which I shall treat under the heading of distributive justice,[5] proves to be a constraint on preference-based choice which is based on the structure of some of the situations in which we make choices. This constraint is generated internally, within the theory of rational choice. That a constraint on preference can be justified by an appeal to preferences may appear paradoxical, but I shall endeavour to remove the air of paradox as we proceed. And as the upshot of my argument I shall insist that distributive justice is not problematic in principle; it may be removed from the area of speculative enquiry, and established securely within rational choice. The age-old philosophical problems about the rationality of morality are solved for the case of distributive justice.

But the firm foundation provided for the constraints on preference-based choice required by distributive justice does not extend to those other constraints which are embraced in our traditional conception of morality. This is why our framework of moral concepts is outmoded. We must distinguish those constraints on preferences which can be justified by an appeal to preference itself from other, external constraints. The latter remains, at least for the present, within the area of speculative enquiry. And here the philosophical problems about the rationality of morality press with renewed vigour.

2. My positive aim in this paper is to show you how we bargain our way into that part of morality which constitutes distributive justice. The bargain is based on our preferences; its outcome is an agreement which constrains our preferences; thus paradox is removed. But before doing this, I want to assure you that in my argument, I permit no sleight-of-hand with the conception of preference, and no question-begging assumptions about the conception of rationality.

I greedily take the largest piece of cake, and you reprimand me for not thinking of the others. Now you might claim that deep down, in my heart of hearts, I really do prefer to consider my fellows. You ask me to reflect. How would I feel were I in their shoes — or had I their appetites? And so forth. Humpty-Dumpty supposed that by paying words extra, we could make them mean what we like.[6] If we pay preference extra, perhaps it will line up with morality. The principles of justice will then reflect our real preferences, requiring us to choose what we really, reflectively, deep-down prefer. Humpty-Dumpty might say this. But Humpty-Dumpty is proverbially confused.

My aim is to ground the theory of distributive justice in the theory of rational choice. In doing this, I generate a part of moral theory from a theory which itself raises no moral issues. But if we insist that our real preferences are moral preferences, then the theory of rational choice is converted into a part of moral theory, and the non-moral grounding of distributive justice is sacrificed. Rather than showing how moral considerations of justice can be generated from non-moral considerations of choice, we should be showing how seemingly non-moral considerations of choice are actually morally based. Paying preference extra, to make it mean what we like, turns our starting-point upside down — like Humpty-Dumpty after the fall.

Thus I shall not talk about "real" preferences — except to refer to what we actually and quite straightforwardly prefer. I really prefer the largest piece of cake. But, one might now say, nevertheless I have good reason to consider the preferences of others. Indeed, one might say, I have as much reason to consider their preferences as to consider my own. So what I should rationally choose is, not that

state of affairs which I personally should most prefer, but rather that state of affairs which would best satisfy everyone's preferences. And that choice would, of course, be just.

Here the sleight-of-hand concerns reason. Rational choice, as I characterized it initially, assumes an essentially subjectivistic and instrumental conception of rationality. What is now urged is that this conception is inadequate. What is rational, it is claimed, must be rational for everyone. On the subjective view, this is taken to imply only that if I choose rationally on the basis of my preferences, then you choose rationally on the basis of your preferences. But it may be alleged that if I choose rationally on the basis of any preferences, then you choose rationally on the basis of those same preferences. On this objective view, the basis of rational choice must include everyone's preferences, or no one's, unless there are intrinsic differences among preferences (or preferrers) such that some count, and some do not.

The objectivistic conception of rationality might seem an ally in my attempt to ground the theory of distributive justice in the theory of rational choice. If objectivity requires that choice be based on everyone's preferences, then fairness seems implicit in the requirements of objective reason. But is objectivity correctly conceived as a requirement of, or a part of, rationality? Although I can not consider this question here, I shall say, quite dogmatically, that I find in every defense of the objectivistic conception of rationality a surreptitious, if not explicit, appeal to moral considerations.[7] The theory of objectively rational choice is thus a part of moral theory, and so can not provide a non-moral grounding of distributive justice.

I do not deny that rationality has implications for morality; indeed, I hope to show what those implications are. But I do deny that rationality is a moral conception. And so I can not appeal to an objectivistic account of rationality which itself depends on moral presuppositions, but only to a subjectivistic, instrumental account which is clearly non-moral. A person acts rationally insofar as he or she seeks to maximize expected utility, where utility is a measure of individual preference. I neither need, nor will accept, any stronger premiss.

3. The link which justice provides between morality and rational choice is discovered by reflection on a phenomenon long of concern to economists, but only recently receiving explicit attention from philosophers. The perfectly competitive market, the ideal of economic theory, is frequently marred by the presence of external inefficiencies. Here is a simple example of an inefficiency.

Several factories must each choose a method of waste disposal. Suppose that air is a free good, so that each factory may discharge effluents into the atmosphere without payment or restriction. Each may then find that it minimizes disposal costs by using the atmosphere as a sink for its wastes. But each factory may also suffer from the pollution occasioned by the effluents discharged. Indeed, it may be that the total cost to all factories, of atmospheric pollution caused by their wastes, exceeds the total net benefit in discharging those wastes into the atmosphere, rather than employing the least costly non-polluting method of disposal. The use of the atmosphere as a sink then constitutes an external inefficiency — external, in that each user displaces the costs of pollution onto others, and inefficient, in that the total costs of pollution exceed the total increase in disposal costs which would be required by an alternative non-polluting method of waste disposal. But no factory has any incentive to adopt such an alternative; each correctly minimizes its own costs by discharging its effluents into the atmosphere.

Bargaining Our Way Into Morality: A Do-It-Yourself Primer

An external inefficiency creates a severe problem for rational choice. We may show this by considering an ideal case, in which each person involved in a situation is able to choose his or her course of action in the light of the actions selected by others.[8] Then, if the persons are rational, each will select that course of action which he or she expects will maximize his or her own utility, given the actions selected by the others. Each action will then be a best reply by the agent to the other's actions. If in any situation the action of each is a best reply to the actions of the others, then the set of actions is a best reply set.

In the presence of an external inefficiency, the outcome of any best reply set of actions is sub-optimal, which is to say that there is at least one other outcome possible in the situation which would better satisfy the preferences of every person.[9] Thus rational choice, given an external inefficiency, leads to an outcome which is mutually disadvantageous, in comparison with some other outcome which the persons could achieve if at least some were to choose differently.

In our example, each factory's best reply to the adoption of waste disposal methods by the others, is to discharge its own wastes into the atmosphere. But if each were to adopt some non-polluting alternative, then all would benefit. It may therefore seem that there is a straightforward solution to this problem created by the external inefficiency — a cooperative solution based on mutual agreement. All of the factories should agree to the least costly non-polluting method of waste disposal. It may then be urged that each factory's true best reply to the others consists in such mutual agreement, and since its outcome is optimal, the inefficiency disappears and there is no problem for rational choice.

Alas, matters are not so simple and straightforward. First, although a non-polluting method of waste disposal reduces total net costs, yet each factory need not benefit. Some factories may suffer greatly from the pollution caused by others, or may find some non-polluting method of waste disposal only slightly more costly than using the atmosphere as a sink, but other factories may suffer very little from pollution, or may find the increased costs of any alternative disposal method very great. Thus an agreement to adopt a non-polluting method of waste disposal, although beneficial on balance, may increase net costs for some factories. To avoid this, the agreement must provide for transfer payments, from those factories which would otherwise benefit most from non-pollution to those which would otherwise not benefit at all. But the amount of compensation is not easily determined. In general, many possible arrangements will leave each factory better off than if all pollute, so that reaching a specific agreement, which each would rationally choose, raises difficulties not only in practice, but for the theory of choice.

Furthermore, although the outcome of an agreement not to pollute may be optimal, and although the outcome of an agreement which includes transfer payments may be mutually advantageous, yet adherence to any agreement need not be the best reply course of action for any factory. Each factory would most prefer that all others cease using the atmosphere as a sink, while it continues polluting. Hence each will be tempted to defect from any agreement, however beneficial the agreement may be. Adherence to an agreement not to pollute, and to compensate any who would not otherwise benefit, is not, in the absence of penalties for violation, the most preferred course of action for any factory, whether the other factories adhere to the agreement, or violate it. Mutual violation thus makes up the best reply set of actions.

External inefficiencies thus raise two problems for rational choice. First, how are we to formulate a specific, optimal, mutually advantageous agreement, or mode of cooperation, for overcoming an inefficiency, which each person affected

will consider it rational to accept? Second, how are we to ensure that rational persons will comply with an agreement so formulated and accepted? These problems may be related, in that we may suppose that compliance with an agreement is rational if acceptance of the agreement is also rational. But this is not evident, and I shall return to the problem of rational compliance in section 6.

4. Let us now focus on the problem of formulating a rational agreement. An agreement consists of a set of actions, one for each person party to it. I assume for the present that compliance is assured, so that no restriction to best reply sets of actions is involved. Now we may say that an agreement takes effect if and only if each party selects the same set of actions. Hence we may represent the problem of formulating agreement as a problem of rational choice — the problem of choosing among alternative possible states of affairs, each the outcome of a set of actions, one for each person involved, subject to the condition that the choice takes effect only if all parties select the same alternative.

This problem arises for anyone who may find him or herself in situations so structured that external inefficiencies arise, or in other words, so structured that no best reply set of actions is optimal. Although not all situations involving interaction among persons have this structure, there can be no assurance against finding oneself in such situations, as long as each individual's preference orderings among alternative possibilities are independent of the orderings of others. So this is a general problem which we all face. Its resolution is not to be found in the particular circumstances in which an individual finds him or herself. Rather its answer must be a general policy applicable to all such circumstances — and, obviously, applicable to all individuals. The policy which any person should adopt, who seeks to cooperate with his or her fellows in the face of external inefficiencies, is and must be identical with the policy every other person should adopt. The content of an agreed set of actions will of course vary with persons, their capacities, preferences, and circumstances, but the form which their agreement takes will be perfectly general.

Consider then the reasoning of a supposedly rational agent — myself — faced with this problem of rational choice. Given an external inefficiency, I must be willing to enter into some agreement with my fellows. Its expected utility to me must exceed the expected utility of failing to agree, which is the utility of my best reply to the actions I should expect others to perform in the absence of agreement. Its expected utility cannot exceed the greatest utility which would be compatible with others receiving only minimally more than they would in the absence of agreement. Thus a utility range is defined, with its lowest point the utility of no agreement, and its highest point the maximum utility compatible with others receiving their "no-agreement" utility. Each person will define such a utility range for him or herself, and only sets of actions which assure everyone a utility within his or her range will be candidates for agreement.

In choosing among candidates some compromise will be required. I must recognize that I am involved in a bargaining situation, and must make some concession.[10] How do I decide the magnitude of the concession which my agreement to some set of actions would require? The answer is implicit in the conception of a utility range. The lowest point of my range represents my point of total concession, in which I gain nothing from agreement. The highest point represents no concession, in which I gain everything. Any intermediate point may be represented as a proportion of my total concession. Not only will this measure my concession; it will relate it to the concessions of others. Two persons make equal concessions

in a situation if and only if each concedes the same proportion of his or her total concession.

Each set of actions which is a candidate for agreement may be represented also as a set of concessions, one for each person. Each such set must have a largest member — the maximum concession required for agreement to be reached on that set. Some possible set of concessions must have a largest member which is no greater than the largest member of any alternative set. This is the minimax concession — the smallest, or minimum, among all possible largest, or maximum concessions.

If there is to be agreement, then someone must make a concession at least equal to the minimax. Now if it is not rational for me to make such a concession, then, since the policy which is rational for me is rational for everyone, it is not rational for any person to make such a concession, and there can be no rational agreement. But it is rational for me to enter into an agreement; hence it must be rational for me to make a minimax concession. Furthermore, since agreement can be reached without any person making a larger concession, and since it cannot be rational for me to make a greater concession than necessary, it cannot be rational for me to make a concession larger than the minimax. Hence it is rational for me to enter into any agreement requiring at most the minimax concession from me. Since everyone reasons similarly, bargaining among rational persons proceeds on the principle of minimax concession. And this solves the problem of rational choice occasioned by external inefficiencies.

We have now characterized a rational bargain. I must next argue that the principle of minimax concession captures our conception of distributive justice, in characterizing a bargain which is fair as well as rational. And I must also argue that the principle constitutes a constraint on preference-based choice, even though it is, as I have shown, itself the outcome of a preference-based choice. Thus I must show that in acting on the principle of minimax concession, we enter into bargains which are fair, and which constrain preference — or in other words, we bargain our way into morally binding arrangements.

One word of warning is in place before proceeding. Although we may literally bargain our way into moral constraints in some contexts, references to bargains and agreements are to be understood hypothetically. We face externalities and, if we are rational, we cooperate to overcome them. We may then assess our mode of cooperation as if it were the outcome of a bargain. But we need suppose no actual bargain or agreement.

5. Under what conditions is a state of affairs distributively just? The presence of more than one person (or perhaps of more than one sentient being) gives rise to a "distribution" of utilities, but this is not sufficient to raise issues of justice. If a state of affairs is said to be just or unjust, there must be at least one alternative to it, the variation in the utility-levels of different persons among alternatives must be at least partially interdependent, and the selection among alternatives must be at least partially a matter of human choice. These conditions are required if any comparison of the utilities received by different persons is to have moral significance. For distributive justice to have significance, distributive considerations must be relevant to the choice among the alternatives. If that choice is adequately represented by a best reply set of actions, then although the choice has distributive effects, these are of no concern to the choosers. It is, therefore, only when all best reply sets lead to sub-optimal outcomes, so that there are mutual advantages to be found in agreement or cooperation among persons, that considerations of distributive jus-

tice arise. Other moral considerations may arise in other contexts, but in restricting distributive justice to the context of mutually advantageous cooperation, we are following in the footsteps of Hobbes, Hume, and Rawls.[11]

This restriction on the scope of considerations of distributive justice suggests that a state of affairs is just, if and only if those involved in it would justly have agreed to the set of actions bringing it about. We must make any reference to agreement hypothetical, since as I have pointed out, much of our social interaction which is at least partially cooperative involves no actual agreement or bargain. But we may replace our question about the justice of states of affairs by one about the justice of agreements, provided we recognize that "Would we agree ...?" rather than "Did we agree ...?" is the appropriate way to introduce reference to such agreements.

The justice of an agreement may be supposed to have two dimensions — one concerning the manner of agreement, the other concerning the matter or content of agreement. But we cannot strictly distinguish these dimensions since in the case of hypothetical agreement, manner reduces to matter. We might say that, ceteris paribus, an agreement is just in manner if and only if it is genuinely voluntary. But the nearest approximation to what is voluntary in the case of hypothetical agreement, must be what is rationally acceptable. And so rationality and justice are inextricably intertwined in our account.

But we may still reflect on the matter of agreement. And here, although rationality and justice are still intertwined, the connection is less direct. For we may say, quite without reference to rationality, that a non-optimal agreement, in depriving someone of benefit unnecessarily, without gain to anyone else, is unfair to the person so deprived. It is unfair for me to be allowed to profit at another's expense, no doubt, but it is equally unfair to me not to be allowed to profit, if no one is worsened thereby. Thus optimality is a requirement of fairness, and so of justice, as well as a requirement of rationallty.

And this is not all. It is unfair to profit at another's expense. How is this unfairness expressed in the context of agreement? Each person's utility range represents his or her possible gain. The expected utility of any proposed agreement may be represented as a proportion of that gain, and so represented, constitutes the relative advantage of the agreement to the person. Now one profits at another's expense insofar as one's own relative advantage can arise only if he or she accepts, not merely a lesser relative advantage, but one less than anyone need accept. Thus one would arrive at a fair agreement by maximizing the minimum relative advantage received by anyone. But the measure of relative advantage is such that for any agreement, the sum of one's relative advantage and one's concession equal unity. Thus maximin relative advantage is equivalent to minimax concession. And so the requirements of fairness and rationality coincide. A hypothetical agreement which is just in manner and fair in matter ls a rational agreement.

The justice of an agreement has been characterized relatively to the set of possible agreements. In other words, a state of affairs is distributively just (or unjust) in relation to alternatives. The set of possible agreements is itself defined relatively to the expected outcome of no agreement. Thus the justice or injustice of a state of affairs is determined against a baseline which provides a certain expected utility to each person, but which itself is not characterized as just or unjust. Any assessment, either of the range of possibilities, or of the baseline, falls outside the scope of considerations of distributive justice, except insofar as the assessment refers to other cooperative arrangements treated ln terms of hypothetical agreement. Such assessment thus constitutes part of the realm of speculative en-

quiry from which distributive justice is freed by its identification with rational choice.

6. Why does a rational bargain, or a mode of cooperation which could be rationalized in terms of a bargain, involve a moral constraint on action? An objector might plausibly argue that insofar as the point of a bargain is to benefit all parties to it, morality has no place. Agreement and cooperation simply constitute an extension of rational prudence.

The apparent strength of this objection rests on ignoring the problem of compliance. This problem has received attention from earlier theorists of justice; although my concern here is not to discuss texts, a quotation from Hume may be illuminating. Hume, I should note, holds a general view of morality strongly opposed to the one I have assumed; he supposes it to further, rather than to constrain, each individual's pursuit of his own interests.[12] But on this view he finds that justice presents a problem:

"Treating vice with the greatest candour, ... there is not, in any instance, the smallest pretext for giving it the preference above virtue, with a view of self-interest; except, perhaps, in the case of justice, where a man, taking things in a certain light, may often seem to be a loser by his integrity. ...a sensible knave, in particular incidents, may think that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy. That honesty is the best policy, may be a good general rule, but is liable to many exceptions; and he, it may perhaps be thought, conducts himself with most wisdom, who observes the general rule, and takes advantage of all the exceptions."

"I must confess that, if a man think that this reasoning much requires an answer, it would be a little difficult to find any which will to him appear satisfactory and convincing."[13]

Hume states the problem of compliance very clearly. Grant that it is rational — or, to use his terminology, preferred with a view to self-interest — to agree on a particular mode of cooperation in situations in which otherwise external inefficiencies would prevent an optimal outcome. Grant that one should adhere to such agreements as a general rule, so that one avoids penalties, maintains one's reputation, and sets others a good example. Yet it is nevertheless advantageous to act on whatever opportunities will prove maximally profitable to oneself, including opportunities to violate one's agreements. And so it is in some cases rational to violate agreements, even though it is unjust.

I reject this conclusion. Adherence to one's agreements does indeed in some situations constitute a genuine constraint on preference-based choice. Were this not so, adherence would not be morally significant. But it is not contrary to reason to adhere, insofar as one is adhering to what is or would be a rational bargain. If one is to overcome inefficiencies by bargaining, then one must be able to expect everyone to adhere to the bargained outcome. It is advantageous to overcome inefficiencies, advantageous to do this by bargaining, advantageous therefore to be able to expect adherence to the outcome, and so, I maintain, rational to adhere to the outcome. Rationality is transmitted from making an agreement, to keeping the agreement.

Elsewhere I discuss this matter at greater length, arguing that the conclusion I have just reached requires a modification in the maximizing conception of rationality — a modification which, however, it is rational to choose.[14] Thus rationality and morality are brought into harmony. Adherence to a rational bargain, one

resting on the principle of minimax concession, is just, and justice is both a requirement of reason, rightly understood, and an imperative of morality, constraining our preference-based choices.

The principle of minimax concession is thus both the object of rational choice for any person faced with external inefficiencies, and a ground of moral constraint. Characterizing all rational bargains and all modes of rational cooperation, it may itself be conceived as the outcome of a meta-bargain — of a supreme hypothetical agreement among all human beings who must interact in situations in which best reply sets of actions are sub-optimal. In accepting the principle of minimax concession, we bargain our way, not into particular moral arrangements, but into morality itself — or at least, into that part of morality constituted by distributive justice.

7. The principle of minimax concession is applied against a baseline situation, and a range of possibilities which must each be mutually advantageous in relation to that baseline. In effect, both the characteristics and the existing circumstances of the persons involved are taken for granted; they provide a framework which determines whether the principle of justice has any application. As Hume noted, the relation between human beings and other creatures who, though rational, lack power to express effectively any resentment against human behaviour, does not involve the restraints of justice. Humans may act as they will, and "as no inconvenience ever results from the exercise of a power, so firmly established in nature, the restraints of justice and property, being totally useless, would never have place in so unequal a confederacy."[15] Hume insisted that animals in relation to humans, barbarous Indians in relation to civilized Europeans, and in many nations the female sex in relation to the male, are in a position of inferiority such that questions of justice and injustice simply do not arise.

Hobbes, who saw in morality a rational response to the horrendous external inefficiencies of the state of nature, and Rawls, who supposed the principles of justice to be the objects of rational choice in circumstances "under which human cooperation is both possible and necessary," have both insisted that one must reason from an initial situation of equality.[16] But this is no part of the present account — or of Hume's theory. Human beings are equally rational, and so all must choose the same principle to regulate their interaction. The worry that one might tailor principles to his or her particular advantage can be seen to be unfounded, once the formal constraints on choice are properly understood. The real worry is that the principle applies to whatever situations do arise, so that, although we bargain our way into moral constraints, we do so from a purely amoral stance. When we eliminate from our account all factors which do not fall within the domain of rational choice — when we eliminate, for example, either Rawls' specially favoured or Hobbes' specially disfavoured no agreement point — we find that distributive justice is an extremely weak constraint on preference-based choice.

An example — quite fictitious, of course — will help to clarify my point. Suppose a planet, the land mass of which consists of two large islands, widely separated by stormy seas. On each, human life — or life close enough to human for our purposes — has developed in complete independence and ignorance of the other. On one island, the Purple People have developed an ideally just society. Knowing the extent of their natural resources, they have adopted policies governing population, conservation, and development, to ensure, as far as they are able, that the worst-off person shall benefit, relative to his or her personal characteristics and the possible modes of social cooperation, as much as possible, not only

in the present generation, but throughout their forseeable future. On the other island, the Green People live in totally chaotic squalor. Taking no thought for the morrow, they have propagated their kind and squandered their resources so that they are on the brink of catastrophic collapse. At this point in their respective histories, an exploration party from the Purple People discovers the Green People, and reports back on their condition.

Consensus among the Purple People is reached on the following points. First, any contact between Purple and Green will require Purple's initiative, since the Greens lack means of both transportation and communication across the ocean. Second, the combined resources of the two islands cannot support the combined populations at the level achieved by Purple society. Third, maximization of the average absolute level of planetary well-being would require a massive but technologically feasible transfer of resources from Purples to Greens. And fourth, the Purple People have the capacity to eliminate the Green People, without any possibility of significant retaliation.

Four parties develop among the Purple People. The first, whom I shall call Utilitarians, demand that the Purples give up their comfortable way of life to rescue the Greens from impending catastrophe and maximize overall well-being. The second group propose that existing levels of well-being in the two societies be taken as a baseline, and the possibilities of mutually advantageous interchange be explored, in line with the principle of minimax concession.[17] This policy, members of the group urge, will maximize the minimum gain relative to existing circumstances, and so will be just. The third group argue that the strains of the continuing inequality between Purple and Greens envisaged in the policy proposed by the second group will outweigh any advantages from interchanges, and urge therefore that no contact be established with the Greens. Finally, the fourth group, whom I shall call Hobbists, argue that the others mistakenly identify the baseline with the existing situation rather than with the outcome of no agreement. Whatever the Greens may seek to do, the best action for the Purples is to eliminate the Greens and appropriate their resources. There is no place for mutually advantageous agreement, and so for consideration of justice.

Let us reflect on these proposals. In my view, many existing moral theories accept far too strong constraints on the maximization of individual utility. Advocates of such theories would find themselves committed to the individually and socially sacrificial policies of the Utilitarian Purples. But not one of us acts on the counterpart of such policies. It is, however, a long step between supposing that one would be literally mad if one took utilitarianism seriously in practice, and supposing that we should accept only that part of morality which can be salvaged with our theory of distributive justice. For we should then be committed to the annihilative policies of the Hobbist Purples, since they recognize that the Purples have no reason to cooperate in any way with the Greens, but rather every reason to eliminate them and acquire their resources.

Of course, it is possible that humanitarian feelings would not only hold the Purple People back from the Hobbist policy, but would make that policy actually less satisfying than one of the alternatives. But surely we should want to say that it would be wrong for the Purples to annihilate the Greens, even if the Purples take no interest whatsoever in the Greens' interests, or feel no emotional concern at all. The Greens, we might even say, have rights, which would be violated were the Purples to annihilate them.[18] There are moral constraints which the Purples should recognize, stronger than any which are generated by mutual advantage.

Either the Purple People should cooperate with the Greens, taking their

present situations as the baseline, or they should leave them alone. Which they should do depends, in my view, on empirical, psychological considerations about the strains of a continuing, unequal relationship. This is an issue in moral psychology, but not directly in moral philosophy. But to defend this position, I require something akin to Nozick's well-known Lockean proviso, as a constraint on the baseline from which mutual advantage is to be determined.[19] In the absence of such a constraint, I see no defense against the Hobbist who insists that the inequality in power between Purples and Greens makes any moral relationship, any moral constraint, irrational.

Thus I come to both an optimistic and a pessimistic conclusion. The optimistic conclusion is that the argument which I have presented grounds a part, and a not unimportant part, of traditional morality, on a strictly rational footing. Using only the weak conceptions of value as individual preference-satisfaction, and of rationality as maximizing preference-satisfaction, I have established the rationality of distributive justice, as that constraint on preference-based choice required by minimax concession.

The pessimistic conclusion is that no similar argument will put the remainder, or any important part of the remainder, of traditional morality on a similarly rational footing. I have not shown this, but we may easily see that the only constraints on preference-based choice which are compatible with our conceptions of value and reason must be those which it is mutually advantageous for us to accept, and these are simply the constraints required by minimax concession. Having abandoned all religious or metaphysical props for morality, we are left with no justification for principles some of which, at least, we are unwilling to abandon.

Related to these conclusions are two opposed views of our society. The optimistic view is that modern Western society is, so far, unique in its recognition that the sole purposes for which coercive authority is justified among human beings are, first, to overcome the force and fraud which are the great external inefficiencies in the state of nature, thus making possible the emergence of the free, competitive market, and second, to assure the efficacy of those modes of cooperation which are required to avoid those public bads and attain those public goods which the free activity of the market will not provide. Until corrupted by the utilitarian and egalitarian ideas which have led to the welfare state, our society was beginning, for the first time in human history, to make it possible for human affairs to be guided by reason and justice.

The pessimistic view is that modern Western society has abandoned every justification for coercive authority and for constraints on preference-based choice save that which stems from consideration of mutual advantage, thereby opening the way to the dissolution of all those genuinely social bonds among human beings which are the necessary cement of any viable public order. That there is a rational resolution of the problem of compliance is of little concern to human beings for whom reason is the slave of the passions, and who, freed from traditional constraints, face a rapid decline into the state of nature conceived as the war of every person against every person.[20]

There is a schizophrenia in these conclusions which I find haunting the core of my moral and political theory. Perhaps we exceed both our hopes and our fears in bargaining our way into morality.

Bargaining Our Way Into Morality: A Do-It-Yourself Primer

FOOTNOTES

[1]John Rawls, A Theory of Justice, Cambridge, Mass., 1971, p. 16.

[2]For further discussion of preference and utility, see R.D. Luce & Howard Raiffa, Games and Decisions, Wiley, New York, 1957, Ch. 2.

[3]Cf. Aristotle, Ethica Nicomachea, 1129b25ff.

[4]For some of my disagreements with Rawls, see my paper "Justice and Natural Endowment: Toward a Critique of Rawls' Ideological Framework," Social Theory and Practice, 3, 1974, pp. 3-26.

[5]Why distributive justice? Because my concern is with justice in contexts in which a distribution of benefits and costs is part of the object of choice. In my view, distributive justice contrasts with acquistive justice; the first constrains modes of cooperation, the second constrains the baseline from which cooperation proceeds. The Lockean proviso (see n.19 infra.) concerns acquisitive justice, and falls in the realm of speculation which is not my concern in this paper.

[6]Cf. Lewis Carroll, Through the Looking-Glass, Ch. VI. "When I make a word do a lot of work like that," said Humpty Dumpty, "I always pay it extra."

[7]The test case here would be the discussion of objective reasons in Thomas Nagel, The Possibility of Altruism, Oxford, 1970, Chs. X,XII.

[8]It may seem that if each person is to choose his or her course of action in the light of the actions chosen by the others, a regress is involved. But in fact the requirement may be operationalized quite straightforwardly. Suppose that each person were to announce his or her proposed action to the others, that after each announcement any other person might announce a new or changed proposal, and that no one were to act until, everyone having made some proposal, no one had announced any change.

[9]Strictly, an outcome is sub-optimal if there is at least one other possible outcome which would better satisfy the preferences of some persons without lessening the satisfaction of any.

[10]Discussions of bargaining theory may be found in J.F. Nash, "The Bargaining Problem," Econometrica, 18, 1950, pp. 155-162; E. Kalai & M. Smorodinsky, "Other Solutions to Nash's Bargaining Problem," Econometrica, 43, 1975, pp. 513-518; and my paper "The Social Contract: Individual Decision or Collective Bargain?," in C.A. Hooker, J.J. Leach, and E.F. McClennen (eds.), Foundations and Applications of Decision Theory, Vol. II, Reidel, Dordrecht & Boston, 1978, pp. 47-67. The account I provide here of a rational bargain parallels a solution offered by Kalai and Smorodinsky.

[11]Cf. discussions in Thomas Hobbes, Leviathan, Ch. 15; David Hume, An Enquiry Concerning the Principles of Morals, Sec. III, Pt.I; John Rawls, A Theory of Justice, pp. 126-130.

[12]"...what theory of morals can ever serve any useful purpose, unless it can show, by a particular detail, that all the duties which it recommends, are also the true interest of each individual?" An Enquiry Concerning...Morals, Sec. IX, Pt.II.

[13]An Enquiry Concerning...Morals, Sec. IX, Pt. II.

[14]Cf. my paper "Reason and Maximization," Canadian Journal of Philosophy, 4, 1975, especially pp. 426-430.

[15]An Enquiry Concerning...Morals, Sec. III, Pt. I.

FOOTNOTES

[16]"If Nature... have made men equal, that equalitie is to be acknowledged: or if Nature have made men unequal; yet because men think that themselves equal, will not enter into conditions of Peace, but upon Equal terms, such equalitie must be admitted." Leviathan, Ch. 15. "It seems reasonable to suppose that the parties in the original position are equal." A Theory of Justice, p. 19. The words quoted in the text are from p. 136.

[17]I shall leave the second and third parties unnamed. I do however believe that the second party could fairly be called Humean, but I cannot defend this claim here.

[18]The reader may (should) be reminded of: "Individuals have rights, and there are things no person or group may do to them (without violating their rights)." Robert Nozick, Anarchy, State and Utopia, Basic Books, New York, 1974, p. ix.

[19]"Locke's proviso... is meant to ensure that the situation of others is not worsened." Anarchy, State and Utopia, p. 175. Clearly the Hobbist policy would worsen the situation of the Greens.

[20]Cf. my paper "The Social Contract as Ideology," Philosophy and Public Affairs, 6, 1977, especially pp. 153-164.