The Case of the Noxious Neighbors

1. TWO THEORIES OF JUSTICE

A Treatise on Social Justice is addressed to the question that Plato asked in the *Republic* two and a half thousand years ago: What is justice? The asking of that question by Plato may be said to have inaugurated political philosophy in the Western world. But the question itself is one that arises inevitably in any society whenever its members start to think reflectively about the arrangements within which their lives are lived. Through contact with other societies, people come to realize that social arrangements are not a natural phenomenon but a human creation. And what was made by human beings can be changed by human beings. This realization sets the stage for the emergence of theories of justice. For a theory of justice is a theory about the kinds of social arrangement that can be defended.

   In Plato's time as in ours, the central issue in any theory of justice is the defensibility of unequal relations between people. Like the Athenians, we see all around us in our societies huge inequalities in political power, in social standing, and in the command over economic resources. The degree of inequality on each of these dimensions is different in different societies, and so is the extent to which a high position on one is associated with a high position on the others. South Africa is not easily confused with Scandinavia. Nevertheless, in every society there are those who give orders and those who obey them, those who receive deference and those who give it, those who have more than they can use and those who have less than they need.

Moreover, even if there is some tendency for those with superior personal characteristics to occupy the higher positions, the correlation is manifestly a weak one, and in any case the height of the social pyramid seems out of all proportion to the range of talent and achievement actually found among people. The implication of this (as Hobbes and Hume both recognized) is that if any existing society is to be deemed just as it stands, the defense will have to be indirect. It will have to take the form of an argument that these gross inequalities are inescapable consequences of the operation of social arrangements with advantages such as liberty, security, or prosperity. Whether or not arguments on these lines are valid is one of the key questions that any theory of justice must reach a conclusion about. (The answer I shall give is a very qualified yes.) But whether or not inequalities of the kind I have described can be defended, there can be no doubt that their existence poses the issue of justice inescapably.

At the time when Plato wrote the *Republic*, nobody seriously questioned the idea that the bounds of justice were the bounds of the state. Then, as now, the violation of treaty obligations was denounced as injustice. But the framework within which the domination and exploitation of one society by another took place was not regarded as open to scrutiny on a charge of injustice. The assertion that Plato put into the mouth of Thrasymachus in the *Republic*—that justice is what is to the advantage of the powerful—was rather shocking as a statement about justice among fellow citizens. But it is exactly this same sentiment that Thucydides has the Athenian envoys put forward in their dialogue with the leading citizens of the island of Melos: "the question of justice only enters where there is equal power to enforce it, and . . . the powerful exact what they can, and the weak grant what they must."[1] And although the dialogue itself is, of course, fictitious, it seems to represent well enough the dominant attitude among the Athenians, if we judge by actions rather than words.

Plato's neglect of (or disbelief in) justice beyond state borders has been pretty faithfully followed by subsequent philosophers who have applied themselves to the topic of justice. What is especially noteworthy is the long-continued failure even to consider the justice of the distribution of wealth in the world as a whole. Indeed, to the best of my knowledge the first extended treatment of this topic by a political philosopher dates from as late as 1979.[2]

In earlier times, this neglect of international distribution was more excusable, for two reasons. First, the means of redistribution were fee-

ble. News traveled slowly and supplies even slower. Only two hundred years ago, Adam Smith could write:

> Whatever interest we take in the fortune of those with whom we have no acquaintance or connection, and who are placed altogether out of the sphere of our activity, can produce only anxiety to ourselves, without any manner of advantage to them. To what purpose should we trouble ourselves about the world in the moon? All men, even those at the greatest distance, are no doubt entitled to our good wishes, and our good wishes we naturally give them. But if, notwithstanding, they should be unfortunate, to give ourselves any anxiety upon that account seems to be no part of our duty. That we should be but little interested, therefore, in the fortune of those whom we can neither serve nor hurt, and who are in every respect so very remote from us, seems wisely ordered by Nature; and if it were possible to alter in this respect the original constitution of our frame, we could yet gain nothing by the change.[3]

Now that men can walk on the moon and send back photographs of the earth from space, this all sounds very quaint. Nor is distance nowadays a bar to the ability to help—or harm. If there are no duties to aid the misfortunate at a distance this will require a moral argument. The plea of incapacity will not work.

The second reason for a new salience to the issue of international redistribution is that it is only in the past two hundred years that the processes of unequal economic development have opened up such enormous international disparities. Even a century ago, the standard of living of the average European industrial worker or agricultural laborer—measured in life expectancy, adequacy of diet, quality of housing, hours of work, and so on—was not outstandingly better than that of a moderately prosperous Asian peasant.

Now, however, the degree of economic inequality of the world population taken as a whole is more extreme than that in all but a very few states in Latin America, the distribution of whose wealth almost everyone would agree to be intolerably inequitable. The justice of the international distribution of economic resources cannot therefore be left aside in any general treatment of justice. I shall discuss some of the issues raised by international justice briefly in chapter 5, and then return to discuss the international situation more extensively in the final volume of this work.

In the *Republic*, Plato discussed two main theories of justice. One is his own, a hierarchical notion according to which a just society is one modeled on a well-ordered human soul. For reasons that will become clear in Part III, I totally reject the presuppositions of this theory and

shall say no more about it. However, the theory against which Plato pits his own—the theory that he presumably regarded as the one to beat—continues to be a live option, and is one of the two theories around which *Theories of Justice* is constructed. Like Plato, I shall eventually reject this theory as inadequate, but I hope to give it a better run for its money than it got from Plato. This is in fact far easier to do than it was when Plato wrote, because the theory has been developed so much further. Hobbes and Hume restated it at length in the seventeenth and eighteenth centuries, and in the past thirty years or so, with the advent of game theory and its application to problems of fair division, it has become possible to work with it in a far more sophisticated way than before. This in turn has stimulated philosophers, of whom David Gauthier is the best known, to recast the theories of Hobbes and Hume taking advantage of the technical refinements now available.[4]

The theory in question is the one that is introduced in the *Republic* by Glaucon:

> People say that injustice is by nature good to inflict but evil to suffer. Men taste both of its sides and learn that the evil of suffering it exceeds the good of inflicting it. Those unable to flee the one and take the other therefore decide it pays to make a pact neither to commit nor to suffer injustice. It was here that men began to make laws and covenants, and to call whatever the laws decreed "legal" and "just." This, they say, is both the origin and the essence of justice, a thing midway between the best condition—committing injustice without being punished—and the worst—suffering injustice without getting revenge. Justice is therefore a compromise; it isn't cherished as a good, but honored out of inability to do wrong. A real "man," capable of injustice, would never make a pact with anyone. He'd be insane if he did. That, Socrates, is the popular view of the nature of justice and of the conditions under which it develops.[5]

This idea is introduced shortly after Thrasymachus has withdrawn from the discussion, and is recognizably an offshoot of his view that justice is what is to the advantage of the stronger. It concedes the central point that justice is founded in advantage, but argues that, in the actual conditions of human life, people can expect to advance their interests more effectively through cooperating with other members of their society than through all-out conflict with them. In Hobbes's terms: peace is better for everyone than a war of all against all.

It should be noticed as a feature of this theory of justice that no special motive for behaving justly has to be invoked. Justice is simply rational prudence pursued in contexts where the cooperation (or at least forbearance) of other people is a condition of our being able to

get what we want. Justice is the name we give to the constraints on themselves that rational self-interested people would agree to as the minimum price that has to be paid in order to obtain the cooperation of others.

The alternative to this that I shall be putting forward is less conceptually parsimonious. Followers of this second approach hold that there has to be some reason for behaving justly that is not reducible to even a sophisticated and indirect pursuit of self-interest. It is thus incumbent upon them to explain what the appeal of justice can be, either to human beings in general or at least to those raised under conditions favorable to moral education. I shall give my own answer later (see especially section 35). However, an outline of the answer can be arrived at by considering the function that, on this alternative view, justice is taken to have in human society.

Let us approach this answer by looking at the common ground between the two theories. They share two features. First, they have in common the idea that questions of justice arise when there is a conflict of interest between different people or groups of people. Second, they also share the idea that justice is what everyone could in principle reach a rational agreement on. Both approaches therefore lend themselves to formulation in terms of some kind of social contract, though the contractual apparatus is not essential and in fact both approaches have been developed in noncontractual forms. (I shall have more to say about the relation to contract in chapter 7.)

How then do the two approaches differ? Very schematically, we can locate the difference in this way. Under the first approach the agreement is allowed to reflect the fact that some people have more bargaining power than others. It is bound to do this because it appeals to self-interest as the motive for behaving justly. If the terms of agreement failed to reflect differential bargaining power, those whose power was disproportionate to their share under the agreement would have an incentive to seek to upset it. The second approach, however, is not constrained by the requirement that everyone must find it to his advantage to be just. It can therefore afford the luxury (which it has to pay for, of course, by finding an alternative motivation for behaving justly) of detaching justice from bargaining power.

This gives us the defining characteristic of the second approach, namely, that justice should be the content of an agreement that would be reached by rational people under conditions that do not allow for bargaining power to be translated into advantage. Obviously this is

very vague as it stands. Everything turns on the way in which the conditions of agreement are filled in, and a whole variety of specific theories of justice have been constructed by building up the conditions in different ways. (See especially chapter 9.) What I want to emphasize here is simply that according to the second theory a just state of affairs is one that people can accept not merely in the sense that they cannot reasonably *expect* to get more but in the stronger sense that they cannot reasonably *claim* more.

The motive for behaving justly is, on this view, the desire to act in accordance with principles that could not reasonably be rejected by people seeking an agreement with others under conditions free from morally irrelevant bargaining advantages and disadvantages. I shall postpone until section 35 my discussion of the strength of this motive and the kinds of social situation in which it is most likely to develop. Here all I need to do is emphasize that, on the second approach, we are not bound by the assumption that the answer to the question "Why should I be just?" must appeal to self-interest.

I shall call this second approach "justice as impartiality," in contrast to the first, which I shall call "justice as mutual advantage." The significance of speaking of "justice as impartiality" is that this approach, however it is worked out in detail, entails that people should not look at things from their own point of view alone but seek to find a basis of agreement that is acceptable from all points of view. The general approach, which calls on people to detach themselves from their own contingently given positions and take up a more impartial standpoint is, of course, a product of the Enlightenment, and everyone who follows it acknowledges a debt to Kant. By far the most significant contemporary figure in that tradition is John Rawls, whose monumental *A Theory of Justice*[6] is in my judgment a work of major and enduring significance. Most of Part II will in fact be largely devoted to a critical exposition of and reflection upon certain central themes in his work.

The two approaches have, as I have made clear, been around for quite a long time. However, in the course of the last thirty years or so a good deal more rigor and precision has entered into the analysis. This is partly the result of the already mentioned technical advances in game theory and decision theory and their assimilation into the literature of political philosophy. But what is equally important is simply that a lot of time and effort has gone into working through alternative ways of setting up the problem of justice within the two approaches and arguing about the pros and cons of each. My object in *Theories of Justice* is to

examine this work and in the course of doing so to argue toward some general conclusions of my own.

Part I is concerned with analyses of justice on a small scale—in fact most of the time the smallest scale on which problems of justice can arise at all, namely, two people. In Part II, I shall move on to the larger context and talk abqut justice within societies. I shall try to show that the two theories of justice are both to be found within David Hume's theory of justice and also both to be found within John Rawls's. I hope by looking with a fresh perspective at these two master political philosophers to gain light on the subject of justice in society. Then in Part III, I shall pull together the discussion of small- and large-scale justice in Parts I and II. I shall seek to show how each of the theories of justice is in fact a family of theories, each member of the family defined by the way in which it specifies key components in the theory. I shall use this scheme to locate the solutions discussed earlier, and thereby I hope make clearer what is at issue in the disputes among recent writers on justice.

2. BARGAINING AND ARBITRATION

In this chapter and the next I shall take up the notion of justice in the simplest possible kind of case: a conflict between two parties over the division of some particular scarce resource. This chapter will be confined to justice as mutual advantage. In the next chapter I shall introduce justice as impartiality by looking at the criticisms that have been made of solutions embodying justice as mutual advantage. I think it is as well at this point to emphasize rather than to gloss over the artificiality of any such analyses. We shall be taking the situations of the two parties as given, without any inquiry into the origins of those situations, and we shall be asking what it would be fair for an arbitrator to decide in this one case considered in isolation from all similar ones. Natural objections arise at once. Should we not talk about the justice of social positions before we can sensibly discuss the fairness of particular decisions? And how can we reasonably ignore the fact that conflicts are normally dealt with by rules covering cases of a certain general kind rather than by one-off arbitrations?

If the purposes of the book were purely practical, these objections would be decisive. But if our purpose is first of all to understand as fully as possible the alternative conceptions of justice, there is much to be said for beginning with the simplest cases, and accepting the inherent

artificiality. The rest of the story can then be filled in later. Indeed, the third chapter, which completes Part I, is devoted precisely to asking what happens when we shift our perspective from one in which we look at conflicts between two people on a one-off basis, asking what a fair arbitrated solution would be, and instead think of rules and institutions whose function is to lay out in advance the terms upon which disputes are to be settled. And the relation of small-scale justice as analyzed here and the justice of the institutions that allocate social positions will be taken up extensively when the groundwork has been laid, in chapter 9.

With this by way of preface, let me now turn to justice as mutual advantage in the context of a two-party dispute over a single issue. The idea of justice as mutual advantage is that the just outcome should represent for both parties a gain over what they would have acquired from a continuation of the conflict. This immediately implies that the process of determining a fair outcome has to be split into two parts. The first consists of establishing a nonagreement point: an outcome that the parties will arrive at in the absence of agreement. The second consists of a prescription for moving the parties from there to a point that preserves their relative advantage at the nonagreement point but is in the set of outcomes that are "efficient," meaning that one party cannot be made better off without the other being made worse off. There are, as we shall see, two competing rationales for this prescription for the move from the nonagreement point to one that is efficient in the sense specified. One, which is most fully within the spirit of justice as mutual advantage, says that the move should be made in a way that reflects the relative bargaining power of the parties. The other appeals to an intuitive notion of fairness and claims that a fair division of the "cooperative surplus" is one that divides it equally between the parties. We shall see, however, that many theorists put forward a criterion of equal gain which has the effect of producing the same outcomes as the rationale that appeals to relative bargaining power.

The present section will follow up the first alternative. According to this, the role of an arbitrator is simply to simulate the results of bargaining. It might be asked why there is any point in bringing in an arbitrator in that case. I shall explain later in this section, under "The Uses of Arbitration," where the arbitrator comes in. But there is no point in even raising that question unless we conclude that it makes sense to talk about a bargaining solution—that is to say, an outcome that rational actors, given their respective strategic advantages and disadvantages, ought to reach. Doubts can be, and have been, raised about the pos-

sibility of carrying out this program in any plausible way, and these
doubts are fundamental to the whole enterprise. Of course, it is open to
anyone to object that we do not get fairness by asking what ideally
rational actors would finish up with if they bargained with each other.
But it is not even worth asking that question unless we think it makes
sense to produce a formula and say that this tells us what ideally ration-
al bargainers would finish up with in any given situation. Still less, I
need hardly say, is it worth arguing about the relative merits of alterna-
tive proposals for the formula.

As so often happens, the technical discussion with which economists
and game theorists are most at home has overwhelmed the discussion of
fundamental issues. There is a plethora of competing operationaliza-
tions of relative bargaining power—the world is full of so-called "bar-
gaining solutions"—but there is a dearth of serious discussion of the
very idea of a bargaining solution. Surprisingly, perhaps, philosophers
have been little help here. In fact, they have tended to be more uncritical
than many of the more technically equipped people. As I shall show in
the next chapter, this is a pattern: philosophers tend to show more con-
fidence in the constructs of game theory than do the more sophisticated
game theorists. Thus, David Gauthier, in his *Morals by Agreement*,
devotes one sentence to observing that "whether there are principles
of rational bargaining with . . . context-free universality of applica-
tion . . . has been questioned." He then goes on to say that "undaunted
by . . . scepticism" he will set out his own theory and say why he prefers
it to the Nash solution.[7]

I shall in Appendix B refute Gauthier's criticism of the Nash solution
and his defense of his own alternative. However, before such questions
are even worth discussing we must first address the notion of a bargain-
ing solution itself. In order to have a definite example of a bargaining
solution to hand, I shall explain the earliest and most popular of such
solutions, the Nash solution. I shall then, using it as my illustration, ask
what can be said in favor of bargaining solutions, what can be said
against them, and what can be said in reply to skeptical attacks. I shall
argue for the realism of bargaining solutions, so long as they are not
made to do too much, and I shall show how the practice of arbitration
naturally lends itself to the use, implicitly if not explicitly, of bargaining
solutions.

Let me begin, then, by setting out as clearly and untechnically as
possible the operation of the Nash solution. Anyone who understands
the workings of the Nash solution can without loss skip the exposition

of it given below under the heading "The Nash Solution," though I hope that what follows in the rest of the section will still be of interest. However, I should emphasize that those who feel they "have the general idea" of the Nash solution are precisely those for whom the exposition is designed. I make so bold as to maintain that much of the discussion of bargaining solutions (such as Nash's) by philosophers has been vitiated by their neglecting to obtain an intimate acquaintance with the work-ings of these solution concepts. A clear sign of this is the tendency of philosophers to take over an interpretation from some game theorist of what it is about a solution concept that makes it come out the way it does and then quote it from one another without examining it for them-selves. The interpretation thus becomes like a parcel that is passed from hand to hand and never unwrapped. I shall seek to substantiate this in relation to the notion of a "threat advantage" in the next chapter (see section 8).

THE NASH SOLUTION

For our purposes, there are two key dates in the analysis of fair division between two people in terms of bargaining. These are 1950 and 1955. In 1950, the mathematician J. F. Nash published an article in *Eco-nometrica* entitled "The Bargaining Problem."[8] Five years later, R. B. Braithwaite, a philosopher at Cambridge University who had worked in philosophy of science and decision theory, was elected to the chair of moral philosophy and published, as his first (and, as far as I know, last) contribution to the subject of his chair, an inaugural lecture entitled *Theory of Games as a Tool for the Moral Philosopher.*[9] In this section I shall say something about the context and significance of Nash's article and explain his solution. In the following section I shall do the same for the brief monograph that Braithwaite based on his lecture.

Until Nash came along, the standard view among game theorists and economists was that bargaining problems had no determinate solution. Thus, John von Neumann and Oskar Morgenstern, in their pioneering work in game theory, *The Theory of Games and Economic Behavior,*[10] maintained that it is possible to say only two things about rational bargaining: first, that, if the parties are rational, neither will accept an agreement giving it less than it could obtain in the absence of agree-ment; and second, that the parties will not reach an agreement such that there is an alternative agreement available under which one would be able to do better without the other doing worse. This corresponded

exactly to the standard economic view according to which we can say that rational trading partners will reach the contract curve but we cannot say where on it they will finish up. Subject to those restrictions, the outcome could be anywhere: its location was held to depend on the "psychology of the parties."

In the usual terminology, all we are supposed to be able to say in general is that the outcome of bargaining will, if the parties are rational, lie on that portion of the Pareto frontier that is above the nonagreement point for both parties. I can explain the notion of the Pareto frontier by saying that it is the set of Pareto-optimal points, and that a Pareto optimum is an outcome such that it is not possible to move away from it in a way that makes one party better off without making another worse off.[11] Thus, take the simplest possible kind of case, in which two people can share $100 in any way that they can agree upon. If they fail to agree neither gets anything. Then the requirement that any agreement must be better for each than the nonagreement point has here the trivial implication that neither will actually hand over money to the other. (The Pareto frontier may well include such transfers.) And the requirement that the outcome should be on the Pareto frontier rather than inside it simply entails that they will agree to divide the whole of the money between them rather than, say, only $90 of it.

It is a good question what the meaning of "rationality" is in this context but one that I shall have to be brief and dogmatic about. I think it is clearly a normative or prescriptive concept rather than a positive or descriptive one. It is, however, a minimally prescriptive or normative concept in that it attempts to deduce the implications of the efficient pursuit of utility. Thus, it would be irrational individually to accept as the outcome of bargaining less than the nonagreement utility because that would be a gratuitous loss of utility; and it would be jointly irrational for the parties to settle for an agreement that was suboptimal because they would be giving up an attainable increase in utility.

Common sense, however, revolts against the conclusion that within the limits I have stated the outcome must turn entirely on the personal characteristics of the parties and nothing else. Suppose that one of the parties is very rich and the other very poor. The rich person, let us suppose, will be little affected by how much or how little of the money he gets, whereas the poor one desperately needs a small portion of it but is much less concerned about getting larger amounts. We have an intuitive feeling that the rich person has an advantage in bargaining here that flows from his position. It remains true that a particularly skillful poor

person in this situation might do well against an ineffective rich one. But we feel that there is an objective inequality in bargaining power here that, with rational bargainers, will result in the rich person getting more than half the money.

Nash's solution to the bargaining problem may be seen (and I suggest should be seen) as an attempt to capture this elusive notion of unequal bargaining power formally. For this he needs a measure of utility, and makes use of the measure that von Neumann and Morgenstern constructed making use of hypothetical choices among lotteries. Thus, if the rich person is indifferent between a certain $50 and an equal chance of nothing and $100, we say that, setting no money at 0 utility and $100 at unity, the utility of $50 is 0.5. And if the poor person is indifferent between a certainty of $25 and an equal chance of nothing and $100, we shall say that the utility of $25 is 0.5. That he derives the same utility (within this system of normalization) from $25 as the other does from $50 reflects his relatively greater anxiety to be sure of getting something. The rich man can afford the luxury of accepting a fair gamble.

This method of representing utilities has come in for a good deal of criticism on the grounds, among other things, that it cannot separate out attitude toward risk. But in the present case that is not an objection because, as we shall see later in this section, the best rationale for the Nash solution incorporates a reference to risk in it. According to this, the Nash solution is a point such that the parties are equally averse to risking the nonagreement outcome by holding out for more than the solution offers. Thus, the parties' attitudes to risk constitute the driving force behind the solution.

The actual form of the Nash solution is that rational bargainers will finish up at the point where the product of the utilities of the parties is maximized, when the nonagreement outcome is assigned zero utility to each party. (I shall henceforward assume that the nonagreement payoffs have been set at zero without making it explicit on each occasion.) I should, however, emphasize that the rationale of the Nash solution is not that it is designed to maximize joint efficiency, except in the uncontroversial sense that it gets the parties to the Pareto frontier. (This point will be taken up in Appendix B.) It is not to be seen as a backdoor way of getting utilitarianism without interpersonally comparable utilities. It is intended rather to represent the results of rational agents trying to do the best for themselves individually.

TABLE 1.1.   AN ILLUSTRATION OF THE NASH SOLUTION

Rich		Poor		
Money	Utility	Money	Utility	Product of utilities
$100	1.0	$ 0	0	0
90	0.9	10	0.4	0.36
80	0.8	20	0.6	0.48
70	0.7	30	0.7	0.49
60	0.6	40	0.78	0.468
50	0.5	50	0.85	0.425
40	0.4	60	0.91	0.364
30	0.3	70	0.96	0.288
20	0.2	80	0.98	0.196
10	0.1	90	0.99	0.099
0	0	100	1.0	0

To illustrate the Nash solution at work, I have in Table 1.1 assigned utilities to our rich and poor people at intervals of $10. The highest product of these $10 intervals comes at a 70/30 split in favor of the rich person; the actual maximum is a little more favorable to the rich person, at about 73/27. The same information is represented in Figure 1.1. The Nash solution comes where the rectangle with the largest area can be drawn within the Pareto frontier, and the outcome is again where the rich person gets $73 to the poor person's $27.

I have devoted some time to setting out the Nash solution because the ideas that it incorporates are crucial in understanding later developments. To review them once more, there are four elements that make up the Nash solution. First, there is the nonagreement point as the point from which the gains derived from agreement are to be calculated. Second, there is the Pareto frontier as the set of points from which the solution must be drawn. Third, there is the solution concept itself, which is designed to allocate the gains in moving from the nonagreement point to the Pareto frontier in a way that reflects relative bargaining power. And fourth, there is the assumption that the only information required to operate an adequate solution concept is information about the von Neumann/Morgenstern utilities of the parties. These features are in fact common to almost all the bargaining solutions that have been proposed as variants on Nash: they differ in the way in which they manipulate the utility information to arrive at an outcome.