



Rational Cooperation

David Gauthier

Noûs, Vol. 8, No. 1, Symposia Papers to be Read at the Meeting of the Western Division of the American Philosophical Association in St. Louis, Missouri, April 27-29, 1974. (Mar., 1974), pp. 53-65.

Stable URL:

<http://links.jstor.org/sici?sici=0029-4624%28197403%298%3A1%3C53%3ARC%3E2.0.CO%3B2-Q>

Noûs is currently published by Blackwell Publishing.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/black.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

*Rational Cooperation**

DAVID GAUTHIER

UNIVERSITY OF TORONTO

I

Why cooperate? To achieve mutually advantageous states of affairs. But why is cooperation necessary? To answer this, we must first consider what is involved in cooperation. I propose this account: a number of persons cooperate, or act in a cooperative manner, if and only if each acts in a way determined by their mutual agreement. So characterized, cooperation depends, not on common objectives, but on common principles of action. These common principles are what is necessary to achieve mutual advantage.

Suppose we do not act cooperatively. Each of us acts on a principle determined by himself alone. According to the received view of economists, social scientists, and some philosophers, each acts rationally insofar as he seeks to maximize his utility, where 'utility' is a purely formal term covering whatever goals and values one may have. Let each of us be rational and fully informed about the situation, including the possible actions and the utilities of everyone. Each of us may then form correct expectations about everyone's actions, and each acts to maximize his utility given his expectations. The outcome of such mutually maximizing actions is in equilibrium, or is an equilibrium outcome, i.e., an outcome affording each person a utility at least as great as he could obtain by acting differently, the actions of the others remaining fixed.

The theory of rational noncooperative action thus shows that rational, fully informed persons attain equilibria. But, although in any situation there must be at least one equilibrium outcome (cf. [5]), in some situations no equilibrium is mutually advantageous. First, there is the problem illustrated by the Prisoner's Dilemma.¹ In a situation with the structure shown by this matrix:

| | | |
|-------------------------|--------------------------|---------------------------|
| | <i>Your first action</i> | <i>Your second action</i> |
| <i>My first action</i> | third best for both | my best, your worst |
| <i>My second action</i> | your best, my worst | second best for both |

each of us maximizes his utility by performing his first action, whatever the other does. The outcome is the unique equilibrium, but the outcome of our second actions would be mutually advantageous. The equilibrium outcome is not optimal, an optimal outcome being one which affords each person as great a utility as possible given the utility it affords each other person. If an outcome is optimal, then no possible outcome affords some person greater utility and no person lesser utility. The outcome of our second actions is optimal but not in equilibrium; the outcome of our first actions is in equilibrium but not optimal. We should prefer the optimal outcome, but acting noncooperatively we attain the equilibrium.

Noncooperative action raises a second problem. Suppose we are matching pennies, with payoffs provided by a “bank”. If we both show heads, I win a nickel; if we both show tails, you do; otherwise we both lose a nickel. The situation is:

| | | |
|---------------------|-----------------------|-----------------------|
| | <i>You show heads</i> | <i>You show tails</i> |
| <i>I show heads</i> | 5¢, 0 | −5¢, −5¢ |
| <i>I show tails</i> | −5¢, −5¢ | 0, 5¢ |

Each of us does best to show heads if the other does (showing heads is in equilibrium), but I do much better. Each does best to show tails if the other does, but you do much better. How to choose?

Here, cooperation is needed, not to attain an optimal outcome, for each equilibrium is optimal, but to attain a more acceptable optimum. If each of us acts noncooperatively, neither knows what to do. But if we can agree to flip a penny, and then each show heads if it lands heads, and tails if it lands tails, the expected payoff for each of us will be $2-1/2\text{¢}$. Actually, one of us will win a nickel and the other nothing, but we are assured that neither will lose and that each stands a fair chance to gain. (No doubt in real life the “bank” would consider such cooperation to be collusion, but in our example we ignore its utilities.)

Thus, we require a procedure for cooperation—for determining an agreed way of acting which will secure an optimal outcome, and a *fair* optimal outcome, to each. Such a procedure constitutes rational cooperation.

II

The rational procedure for selecting an optimal outcome has been studied by game-theorists as the bargaining problem. (Cf. [3]: 121–37.) R. B. Braithwaite has proposed such a procedure as characterizing fairness ([1]: 3–6). The procedure I shall develop has been influenced by the work of Frederik Zeuthen ([7]: 111–21), John Nash ([4]), and John Harsanyi ([2]), but differs from their procedures in ways which I shall not explore here.

In most situations there are many optimal outcomes. (In our game, all probability distributions over both showing heads and both showing tails are optimal.) In choosing among these optimal outcomes, the interests of the persons in the situation are opposed; the choice between any two optima must involve direct conflict between the utilities of at least two persons. The selection of an optimal outcome, then, requires some *interpersonal measure* of the *relative advantage* of each outcome to each person.

We may suppose that in any situation each person will set some utility as the minimum which makes cooperation worthwhile for him. If he could not expect this utility from cooperation, he would act independently rather than in an agreed manner. Further, we may suppose that each person will set some utility as his maximum claim, the most he could expect from cooperation. We shall label these utilities u_{mn} and u_{mx} , or the person's *minimal* and *maximal cooperative utilities*.

The relative advantage of any outcome affording the person a utility between these two is now easily measured. If the utility of the outcome is u_x , the relative advantage is: $(u_x - u_{mn}) / (u_{mx} - u_{mn})$. This takes the value 0 if $u_x = u_{mn}$, and 1 if $u_x = u_{mx}$. It is invariant with respect to the choice of unit and zero-point which are the arbitrary features in the measurement of individual utility. Hence, although we may not assume that the numerical utilities of different persons are comparable, we may assume the comparability of relative advantage.

Note that relative advantage is not a measure of comparative

well-being. If utilities are linear with money, and my maximum claim is 5¢ and my minimum 0, and your maximum claim is \$10,000 and your minimum 0, then if I receive 3¢ and you receive \$4,000, you may receive greater well-being, but I receive greater relative advantage, i.e., a greater proportion of possible advantage.

Let us now formulate the basic condition of rational cooperation. First, a rational person will choose the greatest relative advantage compatible with that received by every other person. Second, he will reject a given relative advantage, if no person need receive such a small relative advantage. Third, he will expect any other rational person to reject a given relative advantage, if no person need receive such a small relative advantage. Hence, rational cooperation must secure an outcome affording the highest minimum relative advantage possible, or *maximin* relative advantage. If there are two or more such outcomes, then rational cooperation must secure an outcome, among those with maximin relative advantage, which affords the highest second minimum. If there is more than one such outcome, the procedure is repeated for the third minimum, and so on. If more than one outcome satisfies this iterated procedure—such outcomes will differ only in the permutation of relative advantage among the persons—then our procedure provides no way of selecting among them; here, I shall assume this problem does not arise. Thus, the condition of rational cooperation is: *cooperation is rational if and only if the outcome of cooperative action affords iterated maximin relative advantage.*

Rational cooperation will afford equal relative advantage to each person, if we assume that persons can randomize collectively over their possible actions. On this assumption, if we plot the possible outcomes in utility-space, they fall on and within a closed convex figure, the outcome-space. The upper right bound of this figure represents the optimal outcomes. The point representing the outcome affording each person his minimal cooperative utility must fall within the outcome-space, for that outcome which makes cooperation minimally worthwhile for each person must be attainable. The point representing the outcome affording each person his maximal cooperative utility must fall outside and above the outcome-space or, exceptionally, on its upper right bound, for the maximum *claims* of rational persons are not in general mutually compatible and, if compatible, can not be met by any non-optimal outcome.

The line joining these points is thus the locus of all points affording equal relative advantage to all. It cuts the upper right bound of the outcome-space at that point which represents the outcome affording maximum equal relative advantage. Any other point on the upper right bound must then afford lesser relative advantage to some person, and so can not be the outcome of rational cooperation. Hence, the outcome affording iterated maximin relative advantage is the outcome which affords maximum equal relative advantage. *Cooperation is rational if and only if the outcome of cooperative action affords maximum equal relative advantage.*

III

To complete this sketch of rational cooperation, we must show how the minimal and maximal cooperative utilities are to be selected. The minimal utility is obviously that utility which a person may reasonably expect, should he refuse to cooperate. In any situation, a person can determine the minimum expected utility which can arise from each of his possible actions—the worst he can do, whatever the circumstances and actions of the others. One of his possible actions must then afford him a minimum expected utility at least as great as that afforded him by any other; this is his *maximin utility*. Since a person can guarantee himself the expectation of his maximin utility by his choice of action, whatever others do, his minimal cooperative utility must be no less than his maximin utility.

In some circumstances it may be more than the maximin. Fully informed, rational, noncooperative persons must reach equilibria. It is easily shown that any equilibrium outcome must afford each person at least his maximin utility. But in some situations the equilibria will afford each person more than his maximin. If the problem is to choose among optimal equilibria, it may still be reasonable to take the maximin as the minimal cooperative utility. But if there is only one equilibrium, not itself optimal yet affording each person more than his maximin, it may be reasonable for each to take its utility to him as his minimal cooperative utility. A fuller treatment would require further examination of this issue; here it suffices to insist that the minimal cooperative utility must be at least the maximin utility.

The selection of the maximal cooperative utility follows

from the selection of minimal utilities. A rational person seeks to maximize his utility; he therefore wishes to *minimize* the relative advantage assigned to him for each optimal outcome. Hence, he wishes to set his maximal cooperative utility as high as possible. However, his maximum claim cannot reasonably exceed the greatest utility he might receive, given that each other person receives his minimal cooperative utility. Thus, each person will claim this as his maximum. In effect, each will advance the claim that he alone should receive the benefits of cooperation.

The condition of rational cooperation, stated previously, may now be expanded to include, as an explication of relative advantage, the requirement that it be measured in terms of (i) the minimal cooperative utility for each person, which is at least his maximin utility, and (ii) the maximal cooperative utility for each person, which is the greatest utility he can receive compatibly with each other person receiving his minimal cooperative utility.

IV

Rational cooperators achieve outcomes which are mutually as advantageous as possible and which afford each person an equal measure of relative advantage, or at least afford the least favoured as much relative advantage as possible. Rational cooperation satisfies both a *productive* condition and a *distributive* condition, with respect to human well-being. It may then seem plausible to identify rational cooperation with morality, not in supposing that rational cooperation requires us to do whatever we ordinarily take to be morally right, but rather that it requires us to do what on reflection seems reasonable and justifiable in our moral practices. It enables us to subsume morality under rationality; what cannot be so subsumed we regard as irrational, and cease to consider moral.

In particular, rational cooperation brings into relief those aspects of morality which we associate with justice and fairness. (Cf. [1]: 3–6, 50–55.) The procedure of rational cooperation determines fair outcomes, within the fixed constraints imposed by the structure of situations. Each person is treated fairly if cooperation brings him as great a measure of his well-being, relative to what is possible for him in the situation, as each other person.

The procedure for rational cooperation assures fairness only *within* a situation. It is not a remedial procedure to rectify injustices which are present in the structure of the situation itself. Such injustices, on this view, are not to be discerned simply by attending to the structure; there are no situations which, in virtue of the possible actions and utilities available to the persons in them, are unjust. Rather, situations are unjust insofar as their structure results from previous unjust action. Hence, men who always cooperate in a rational way will be entirely free from injustice in their dealings with each other; the different levels of well-being which they may attain will be the result, not of unfair actions, but of natural circumstances.

In another place it would be interesting to compare the theory of justice which emerges from supposing that rational cooperation ensures fairness, with the superficially somewhat similar theory of John Rawls ([6], esp. Chs. I–III). But before one can develop any such comparison profitably, it is necessary to draw out some of the fundamental implications of the present theory. For it is only in the light of these implications that one may reasonably decide whether it is plausible to present the theory of rational cooperation as a theory of morality and justice.

V

The most significant implications arise from the roles played by the minimal and maximal cooperative utilities. The minimal cooperative utility represents that point at which an individual prefers to act on the basis of mutual agreement rather than to act in a directly maximizing manner, or, we may say, prefers civil society to the state of nature. There is no reason to suppose that this point is the same for all men. To the extent to which one person can expect to do better for himself in the state of nature than another, he will demand a higher minimal return from cooperation. Thus, the minimal cooperative utility reflects what we may call the *natural inequalities* among men.²

Rational cooperation takes these natural inequalities as given. It makes no attempt to alter any balance of advantage which one person can gain over another in the state of nature. In society, each person is entitled to what he could attain for himself, as a result of his natural capacities and his natural

willingness to exert himself, and only after each person receives this, is the remaining well-being to be apportioned by maximizing minimum relative advantage.

Justice, on this view, is not concerned with natural inequalities. Justice is an artificial virtue, the virtue of social practices to overcome the inequalities of nature. Although justice is the fundamental social virtue, it is not a fundamental human virtue. It is only insofar as there are advantages to be attained by cooperative action that justice enters into human affairs. I treat you justly or unjustly, not in interacting with you, but only in cooperating with you. If we are in a zero-sum situation, in which any gain for me is a loss for you and vice versa, considerations of justice or fairness do not arise between us.

Now one can expect objections to this view. Perhaps the basic objection turns on the fact that from the human or social point of view, natural inequalities are entirely arbitrary. (Cf. [6]: 72–75, 102.) Why then should they be maintained? Why should society not seek the equal well-being of all, eliminating or rectifying natural inequalities in so doing?

It must first be emphasized that rational cooperation does not take natural inequality as a basis for further social inequality. I am not entitled to a greater share of the benefits produced by cooperation, if in the state of nature I can expect to do better than you. We must distinguish clearly between apportioning social benefits on an unequal basis, proportional to natural inequality, and apportioning social benefits on an equal basis, after taking natural inequality into account. The present theory requires the latter.

The reply to the objection is that it depends on a view of society incompatible with that presupposed throughout the present argument. Society, as conceived here, is essentially an instrument which individuals mutually accept in order to achieve, for each, benefits unattainable without such a collective instrument. No individual can be expected to accept this instrument if it will not benefit him to do so—if, that is, it will not assure him what he can attain for himself without it, and then add a fair share of those benefits which can be attained only through cooperative action. This is, of course, only an initial reply; what it shows is that the conception of justice in the theory is that appropriate to what our tradition of political philosophy has distinguished as civil society.³

VI

The maximal cooperative utility is the sum of the minimal utility and the maximum potential benefit of cooperation. In other words, the maximal cooperative utility adds, to what the person acting independently can assure himself, the *total* benefit which cooperation secures, insofar as that benefit can be provided to the particular person. To treat people equally and fairly, according to the procedure of rational cooperation, is to provide each person with as great a share of his potential cooperative benefit as each other person, where each share is of course as large as possible.

Just as there is no reason to suppose that the minimal utilities of different persons will be equal, so there is no reason to suppose that their potential cooperative benefits will be equal. If it is possible to make exclusively available, to any person whatsoever, the total benefit which cooperation secures, and if each person values this benefit equally, then the potential cooperative benefit must be the same for each. But on the much more plausible assumption that this is not possible, a cooperative situation will contain different levels of potential benefit for the different cooperators. And since each receives an equal share, measured in terms of potential benefit, a cooperative situation will in fact provide different levels of actual benefit for the different cooperators.

If, as before, we identify cooperative procedures with society, individuals are justly treated if their actual social benefits are equally proportionate to their potential social benefits. Rational cooperation thus preserves what we may call the *fixed social inequalities* among men. These inequalities are built into the very structure of cooperative and social situations, just as are natural inequalities. That there should be any particular fixed social inequalities is, from the point of view of society, an arbitrary matter. But from the points of view of the individual persons who create and maintain society as their collective instrument, these inequalities are not arbitrary. Each individual considers his maximal cooperative utility to represent his social potential. Each recognizes that he cannot reasonably expect the proportion of this potential which society enables him to realize to exceed the proportion of any other person's social potential which society enables that person to realize. But each

may reasonably require that his proportion equal that of any other person.

Thus, the equality which rational cooperation ensures is equality of *opportunity*. Each person has the same degree of opportunity to realize his social potential. To accept any stronger form of equality would be to require some members of society to make proportionately greater sacrifices than others, and this would be rationally unacceptable to utility-maximizing individuals.

VII

In the preceding three sections, I have been characterizing, not morality, but *a morality*—the morality of civil society. This morality prescribes that behavior which is rational for men who satisfy two conditions: they recognize that it is mutually advantageous for them to act on the basis of mutual agreement, and their concern with each other extends only to the effects each can have on the well-being of others. The first of these conditions has been developed explicitly; rational cooperation is the manner of acting which follows from mutual agreement. The second has been assumed implicitly, in identifying rationality with utility-maximization. For it is part of the “received view” of rationality, and hence of man, which I have adopted throughout.

Rational cooperation constitutes the morality of economic men. The relations of economic men one with another are accidental, external; it is no part of their nature, as they conceive it, that each should affect the well-being of others in whatever ways he does. Hence, to them morality appears as artificial and instrumental, as a human contrivance whose only rationale is to assure the greater well-being of each. If as moral men they treat each other as ends, accepting for others the same relative advantage they demand for themselves, yet in a deeper sense they treat each other as means, for the principles of optimization and fairness are to each but part of an overall policy of individual utility-maximization.

For economic men the ideal society is of course the market. The market resolves the problem of enforcement raised by our account of rational cooperation—how do we ensure that each man plays his part in bringing about the fair optimal outcome? In the market situation there is a single, optimal

equilibrium outcome. The existence of such an outcome makes distinctively cooperative action unnecessary. Each person acts to maximize his own utility and thereby cooperates with his fellows to bring about the mutually advantageous, fair outcome. Within the market, the principles of optimization and fairness are the principles of self-interest; morality and prudence are one.

My concern is neither to attack nor to defend the morality of civil society. I do want to argue that *if* you accept the view of man, and of rationality, which is implicit in the identification of rationality with utility-maximization, *then* you must accept this conception of morality, on pain of denying the rationality of morality, and denying it, not just in the weak sense in which moral requirements are held to go beyond rational requirements, but in the strong sense in which some moral requirements are held to contradict rational requirements. Such a denial seems plainly unsatisfactory.

Hence, if the prospect of a world of rational cooperators fails to please, there is no point merely in proposing an alternative moral ideal. What our account of rational cooperation reveals is that the real issues concern what it is to be human, and, being human, what it is to be rational. This study is then intended as a prolegomenon to an exploration of those interconnected conceptions of man, rationality, society, and morality, which constitute our own ideology. Rational cooperation, as the point at which our conceptions of man and rationality are tied to our conceptions of society and morality, is a key nexus in the ideology of economic man and market society.

Appendix

To illustrate the procedure of rational cooperation, consider these examples.

1. A husband and wife are considering whether either (or both) should have an affair. He has a strong desire to do so, but his aversion to his wife having an affair is three times as strong. She has a weak desire to do so, and her aversion to her husband having an affair is one-and-one-half times as great. Since utilities are not interpersonally comparable, the strength of his desires in relation to hers is of no significance. Letting the worst outcome for each have the value 0, we assign

utilities, in accordance with the information above, to produce this matrix:

| | <i>Wife has affair</i> | <i>Wife doesn't</i> |
|---------------------------|------------------------|---------------------|
| <i>Husband has affair</i> | 1, 2 | 4, 0 |
| <i>Husband doesn't</i> | 0, 5 | 3, 3 |

The sole equilibrium results if both have an affair; this assures each his maximin utility, which we take to be the minimum acceptable for cooperation. The maximum claims are then $3\text{--}1/3$ for the husband (since he can receive no more if his wife's expected utility is at least 2) and $4\text{--}1/3$ for the wife. Rational cooperation leads to that optimal outcome with expected utility $2\text{--}4/10$ for the husband and $3\text{--}4/10$ for the wife; he does not have an affair and she randomizes with a 20% probability of having one. His concern about what she does, in relation to what he does, is relatively greater than her concern about what he does, in relation to what she does; this gives the wife the advantage.

2. Braithwaite's example ([1]: 8–11, 21–23): Luke is a classical pianist, Matthew a jazz trumpeter, and each wishes to practice at the same time in adjoining flats constructed without regard to acoustical considerations. They seek a fair division of their practice times. Their utilities are represented in this matrix:

| | <i>Matthew plays</i> | <i>Matthew does not play</i> |
|---------------------------|----------------------|------------------------------|
| <i>Luke plays</i> | 1, 2 | 7, 3 |
| <i>Luke does not play</i> | 4, 10 | 2, 1 |

There are two optimal equilibria, which arise if one plays and the other does not. The maximin utilities, which we take as the minimum acceptable for cooperation, are $3\text{--}1/4$ and $2\text{--}4/5$; the maximum claims are of course 7 and 10; the outcome determined by the procedure for cooperation provides Luke an expected utility of $5\text{--}113/319$ and Matthew an expected utility of $6\text{--}268/319$. It is achieved if both adopt a common randomized strategy affording Luke roughly a 45% chance of playing, and Matthew a 55% chance of playing, on any given night. Matthew's advantage arises because he makes a greater concession in listening to Luke play than Luke makes in listening to him play.

REFERENCES

- [1] Braithwaite, R. B., *Theory of Games as a Tool for the Moral Philosopher* (Cambridge: Cambridge University Press, 1955).
- [2] Harsanyi, John C., "Approaches to the Bargaining Problem Before and After the Theory of Games," *Econometrica* 24 (1956): 144–57.
- [3] Luce, R. D. and Raiffa, H., *Games and Decisions* (New York: Wiley, 1957).
- [4] Nash, J. F., "The Bargaining Problem" *Econometrica* 18 (1950): 155–62.
- [5] ———, "Noncooperative Games," *Annals of Mathematics* 54 (1951): 286–95.
- [6] Rawls, John, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971).
- [7] Zeuthen, Frederik, *Problems of Monopoly and Economic Warfare* (London: G. Routledge & Sons, 1930).

NOTES

*I am grateful to the Canada Council for research support during part of the period in which the ideas in this paper were developed. Versions of this paper have been read to the Vicious Circle at Toronto, and at the University of Sussex, Queen's University (Kingston), and York University (Toronto).

¹A. W. Tucker is the source of the Prisoner's Dilemma.

²The numerical measure of the minimal utility does not itself show natural inequality, for the measure is arbitrary. I am here supposing some way of comparing the utilities of different persons which is not shown by the measure of individual utility.

³Hegel's distinction of civil society from both the family and the state is especially relevant here.