The Economic Journal, **94** (*December* 1984), 772–787
Printed in Great Britain

RECIPROCITY: THE SUPPLY OF PUBLIC GOODS THROUGH VOLUNTARY CONTRIBUTIONS*

Robert Sugden

In modern economies, there are two main ways of financing the production of goods and services. One way is by charging consumers: if you consume, you must pay. The other way is by raising taxes: whether you consume or not, you must pay. But there is also a third way, characteristic of what I shall call the voluntary sector, which is to finance production out of voluntary contributions: whether you consume or not, you choose for yourself whether you pay. There can be no doubt that this third method does sometimes work. In the United Kingdom, for example, the lifeboat service is financed by voluntary contributions and the blood transfusion service is dependent on unpaid donors. Much medical research is funded by gifts; many theatres, orchestras and sports clubs are able to continue only through the success of their fund-raising appeals; trade unions manage to exist where there is no compulsion on anyone to join. This is an economic phenomenon that needs to be explained. In this paper I shall propose a theory of the voluntary sector, based on the assumption that most people believe free riding to be morally wrong.

I. PUBLIC GOODS AND FREE RIDERS

The voluntary sector differs from the profit-making sector in that exclusion is not practised: however the goods and services produced in this sector are allocated among consumers, the fact that a person has contributed towards the costs does not give him any entitlement or priority. Thus one person's contribution typically confers benefits on a *group* of people. In this sense, the services provided by the voluntary sector are public goods, even if – as in the case of blood transfusions and lifeboat rescues – exclusion is technically feasible.

It is remarkably difficult to produce a satisfactory theory to explain how public goods come to be supplied through voluntary activity when many individuals are involved. The problem, of course, is the incentive for each individual to take a free ride.

One obvious way of trying to explain this phenomenon is to assume that each person allocates his income between private consumption and public goods so as to maximise his own utility, taking the behaviour of everyone else as given. In short, we assume 'utility maximisation' and 'Nash conjectures' (cf. Sugden, 1982). It might seem that a theory based on these assumptions can explain the existence of voluntary activity while showing it to be Pareto inefficient. A number

* The first draft of this paper was written while I was enjoying the hospitality of the Center for Study of Public Choice at Virginia Polytechnic Institute. I am grateful to Geoffrey Brennan, James Buchanan and Howard Margolis for comments on earlier versions of the paper.

of writers have argued exactly this in relation to charitable activity – which they take to be a public good because of donors' altruistic concerns about the welfare of recipients (e.g. Schwartz, 1970; Becker, 1974; Collard, 1978, ch. 10; Arrow, 1981). This idea, however, is mistaken. A theory of this kind cannot account for those cases – common enough in reality – in which a public good is paid for by the small contributions of many individuals (cf. Margolis, 1982, pp. 19–21; Sugden, 1982).

This sort of theory has an additional problem. If an individual takes other people's contributions as given, she will contribute *less* as other people contribute *more*. (To derive this result, we need assume only that private consumption is a normal good – that the individual does not reduce her consumption as her income increases.) If many people are contributing to a public good, it turns out that any increase in one person's contribution is almost completely offset by decreases in other people's contributions (Sugden, 1982, p. 346). In the light of this result, the assumption of Nash conjectures seems arbitrary.

It would surely be more reasonable to require conjectures to be *consistent*. Then equilibrium is a state in which each person, maximising her utility on the basis of her conjectures about other people's behaviour, is led to behave in a way that validates their conjectures about her. In such an equilibrium, everyone would expect other people's contributions to be negatively related to her own, and so even less of the public good would be supplied than in a Nash equilibrium. In other words, relaxing the assumption of Nash conjectures makes it *more* difficult to explain voluntary activity in terms of utility maximisation.[1]

If we are interested in the question 'How *might* people overcome the free-rider problem?' there is scope for ingenuity in designing suitable procedures. Guttman (1978) has produced one very interesting solution to this problem – a two-stage procedure, in which individuals first declare their 'matching rates' and then their 'flat contributions'. (The matching rate is the rate at which one person 'matches' other people's flat contributions.) But if we are trying to answer the question 'How *do* people overcome the free-rider problem?' we cannot *invent* procedures; we must work with the procedures that actually operate. Perhaps the lifeboat service *ought* to require its donors to follow the Guttman procedure; but the fact is that it doesn't.

Given the difficulty of explaining voluntary activity in terms of conventional economic theory, it seems worth considering the possibility that individuals act according to some moral principle that requires them to take account of other people's interests.

II. THE PRINCIPLE OF RECIPROCITY

Economists are often tempted to model 'concern for others' by assuming that individuals derive utility from one another's welfare. It is clear, however, that this approach will not provide a solution to the present problem. The assumption of altruistic preferences merely makes each person's welfare into a public good

[1] Cornes and Sandler (1984) claim that it is possible for a non-Nash theory to predict that *more* of the public good would be supplied than would occur at a Nash equilibrium. They can arrive at this result only by implicitly assuming *either* that private consumption is an inferior good (which is extremely implausible) *or* that individuals' conjectures are inconsistent.

from which everyone derives utility, and we are back to square one: why would anyone contribute towards *this* public good? As long as we assume that people maximise utility – however altruistic we suppose their preferences to be – this problem will remain. It seems that if we are to explain why people do not free-ride, we must find a theory that is not based on utility maximisation.

One such theory has been presented by Margolis (1982). This is a general theory of non-selfish behaviour; but Margolis sees one of its main virtues in its ability to explain voluntary contributions to public goods. In this theory each individual has *two* utility functions. One kind of utility – '*S*-utility' – is essentially the individual's self-interest; the other – '*G*-utility' – is the individual's conception of the welfare of the 'group' to which he feels he belongs. The individual allocates his resources between these two departments of life – or two selves – according to some notion of 'fair shares'; then with the resources at their disposal the *S*-self maximises *S*-utility while the *G*-self maximises *G*-utility. Notice that this is a theory of *altruism*: to the extent that the individual acts non-selfishly, he is motivated by a concern for other people's welfare. I shall say more about this theory later.

A very different approach is to suppose that people follow a morality, not of altruism but of *cooperation*. Theories of this kind assume that individuals pursue self-interest subject to moral constraints, and that these constraints are rules which – roughly speaking – it is in everyone's interest that everyone should follow. Several economists have suggested that individuals work on the following rule. Consider any public good whose production involves effort on the part of individuals. Suppose that effort can be measured in a single dimension and that interpersonal comparisons of effort are possible. Suppose that I is the group of individuals who benefit from the public good, and consider any person i who is a member of I. Let i choose the level of effort that she would most prefer that every member of I should make (the same effort for each person). Then i is obliged to make at least this effort. This rule has often been called 'Kantian' (cf. Laffont, 1975 and Collard, 1978 and 1983); Harsanyi (1980) has called it a principle of 'rational commitment'.

Notice that this principle requires each individual to make whatever contribution she would wish others to make – *irrespective of whether the others actually make this contribution.* (For this reason I shall call it the *principle of unconditional commitment.*) Many people, I think, would find the demands of this principle unacceptably strong, and perhaps even morally objectionable. Suppose you have good reason to know that no one else in your group will contribute anything towards a certain public good, irrespective of what you do. The only beneficiaries of your contribution would be yourself and the other members of the group. Why are you obliged to help them, when they refuse to help you?

Perhaps you believe (as I do not) that you *are* morally obliged to contribute in these circumstances. Even if you believe this, you will surely recognise a psychological barrier against contributing: it seems *unfair* that you alone should bear the costs of the public good. And no one else in the group is in a position to urge you to meet your moral obligation; they are not meeting their obligations either. Whatever the force of the principle of unconditional commitment at the level of moral theory, it is hard to see it taking root as a maxim of *practical* morality – as a maxim on which ordinary people are prepared to act.

But suppose instead that everyone else in your group *is* contributing towards a public good from which you benefit: everyone else has paid his union subscription; everyone else is taking his litter home from the beach; everyone else is contributing towards the cost of the office Christmas party to which you intend to go. Now, surely, there is a much stronger moral argument that you ought to contribute, even if it is still not in your self-interest to do so. You also have to reckon with the sense of grievance that the others will almost certainly feel if you refuse to contribute, and with the possibility that they will find ways of punishing you if you do.

It seems, therefore, that a weaker version of the principle of unconditional commitment would be more compatible with most people's sense of practical morality. Roughly speaking, we need a principle that says, not that you must always contribute towards public goods, but that you must not take a free ride when other people are contributing.

I suggest that we reformulate the principle as follows. Let G be *any* group of people of which i is a member. Suppose that every member of G except i is making an effort of at least ξ in the production of some public good. Then let i choose the level of effort that he would most prefer that every member of G should make. If this most preferred level of effort is not less than ξ, then i is under an obligation to the members of G to make an effort of at least ξ. I shall call this the *principle of reciprocity*.

Notice that the principle of reciprocity never requires you to contribute more than other people in the 'group', thus overcoming the objection of unfairness that can be made against the principle of unconditional commitment. However, it is important to notice that *any* collection of individuals can count as a group. The individual has obligations, not to 'society', but to any group of individuals from whose efforts he derives benefits. Groups need not be formally constituted organisations. The groups that have claims on the individual may be occupational, racial, religious or political; they may be local, national or international. 'The set of all people who contribute at least ... to ...' is a group; and so even if some people steadfastly refuse to contribute anything towards a public good, anyone who benefits from the public good has obligations towards those *do* contribute.

Having described the principle of reciprocity in an intuitive way, I shall now show how it can form the basis of a theory of the supply of public goods. I shall assume that everyone accepts the reciprocity principle as a morally binding contraint and investigate the implications of this. Finally I shall compare these implications with those of two other theories – Margolis's theory of altruism, and the theory of unconditional commitment.

III. A MODEL OF THE VOLUNTARY SECTOR

I shall investigate the implications of the reciprocity principle within a very simple – but quite general – model of the voluntary sector. In the model there are n individuals and a single public good. The utility u_i of each individual i is an increasing function of the quantity of the public good, z, and a decreasing function of the *effort*, q_i, that he contributes towards the production of this good:

$$u_i = u_i(q_i, z) \quad (i = 1, \ldots, n). \tag{1}$$

Let $h_i(q_i, z)$ be the marginal rate of substitution (MRS) between z and q_i:

$$h_i(q_i, z) = -\frac{\partial u_i/\partial q_i}{\partial u_i/\partial z} \quad (i = 1, \ldots, n). \tag{2}$$

I shall assume that

$$\partial h_i(q_i, z)/\partial q_i > 0 \quad (i = 1, \ldots, n) \tag{3}$$

and

$$\partial h_i(q_i, z)/\partial z > 0 \quad (i = 1, \ldots, n). \tag{4}$$

These restrictions are natural enough for a utility function defined for one good and one 'bad', and are compatible with various definitions of 'effort'.

I shall leave open the question of how 'effort' should be interpreted, so as to allow the model to accommodate a number of alternative specifications of the reciprocity principle. Notice that there is a normative dimension to the concept of effort. The reciprocity principle says, with certain qualifications, that if everyone else contributes a particular level of effort to the production of a public good, you must do the same. Different definitions of effort lead to different propositions about individuals' obligations.

One possible conception of effort is *effort as labour time*. Suppose that individuals' contributions towards the public good are in the form of labour time, and that effort is measured in terms of hours of labour. Expressions (1)–(4) then amount to the assumption that the individual derives utility from leisure and from the public good, and that these are both normal goods.

Another conception is *effort as absolute money contribution*. Suppose that individuals' contributions are in the form of money, and that effort is measured in money units. Expressions (1)–(4) then amount to the assumption that the individual derives utility from his own consumption of private goods and from the public good, and that these are both normal goods.

This second version of the reciprocity principle might be thought objectionable because it takes no explicit account of ability to pay: there is no obligation on richer people to contribute more than their poorer fellows. A relatively simple alternative is the idea of *effort as relative money contribution*: a person's effort is measured by the size of his money contribution as a proportion of his income.[1] (Compare the practice of tithing to raise money for churches, or the idea that all nations in a military alliance should devote the same share of national income to defence.) Provided that each individual's income is taken as given, this conception of effort is quite compatible with my formal model; all that is being assumed in expressions (1)–(4) is that private consumption and the public good are both normal goods.

I shall assume that the production function for the public good takes the form:

$$z = f(\sum_i \alpha_i q_i). \tag{5}$$

[1] This idea was first suggested to me by Charles Feinstein.

The idea here is that $\sum_i \alpha_i q_i$ measures the total effort of all individuals as an input to the process of production; α_i is a positive constant for each individual i. These constants are necessary in the general model because equal efforts on the part of different individuals need not be equally productive. The function $f(.)$ is assumed to be continuous, increasing and concave, with linearity allowed as a limiting case. This is quite conventional.

Before formulating the reciprocity principle, let me define a further function $F(.,.)$. This function is defined for a given vector of individuals' efforts or *contributions*, $\mathbf{q} = (q_1, ..., q_n)$. For any group of individuals G, and for any level of effort $\xi \geqslant 0$, $F(G, \xi)$ is defined by

$$F(G, \xi) = f(\sum_{j \in G} \alpha_j \xi + \sum_{k \notin G} \alpha_k q_k). \tag{6}$$

Thus $F(G, \xi)$ is the amount of the public good that would have been produced if every member of G had contributed ξ and if each non-member k had contributed q_k. Notice that this function must be continuous, increasing and concave in ξ.

Now consider any group G, and any individual i who is a member of G. Take as given the contributions q_k of all people k who are *not* members of G. Now let q_i^G be the value of ξ that maximises $u_i[\xi, F(G, \xi)]$. If i could choose a single level of contribution for all members of G, this is the level he would choose.

According to the reciprocity principle, i is obliged to contribute at least q_i^G, provided that every other member of G does the same. And if some members of G are contributing less than q_i^G, i is obliged to contribute at least as much as everyone else in the group. This can be put more formally as follows:

> *Obligations*. For any vector of contributions $\mathbf{q}$, for any group of individuals G, and for any member of that group i: i is meeting his obligation to G if and only if *either* (a) $q_i \geqslant q_i^G$ *or* (b) for some other person j in G, $q_i \geqslant q_j$.

Notice that this definition allows 'obligations' to arise in the case where $G = \{i\}$: in this case, i is 'obliged' to contribute at least q_i^G. Clearly, this cannot be an obligation of *reciprocity*; if i is the only member of the group, there can be nothing for him to reciprocate. In this case, q_i^G is the contribution that it is in i's self-interest to make: it is the contribution that maximises his utility, given the contributions of everyone else. Since I shall be assuming that people pursue self-interest within the constraint of the reciprocity principle, it is convenient to say that each person has an obligation to himself to contribute at least as much as self-interest requires. Then if person i makes the smallest contribution that is compatible with his obligations to *all* groups of which he is a member – including the group $\{i\}$ – he must be maximising his utility subject to the moral constraints imposed by the reciprocity principle. An equilibrium state of the model is one in which this is true for all persons i:

> *Equilibrium*. An equilibrium is a vector of contributions $\mathbf{q}$ such that for each person i, given the contributions of everyone else, q_i is the smallest contribution that is compatible with all of i's obligations.

This completes the specification of my model. I shall go on to show that equilibrium exists, and to examine some of the properties of equilibrium, first for for a simple case and then in general.

IV. A SPECIAL CASE: IDENTICAL INDIVIDUALS

Consider the special case in which all n people have identical preferences (defined in (q_i, z) space) and that the production function has the simple linear form $z = \beta \sum_i q_i$. It is obvious from the symmetry of the problem that each person's obligations must be the same as everyone else's. Thus in equilibrium everyone must make the same contribution; the quantity of the public good will thus be given by $z = \beta n q_i$, where i is any individual.

For every individual i, $z = \beta n q_i$ is a ray in (q_i, z) space. Along this ray, $h_i(q_i, z)$, i's MRS between the public good and effort, must increase continuously. Let $(q', \beta n q')$ be the point on the ray at which $h_i(q_i, z) = \beta$, and let $(q'', \beta n q'')$ be the point at which $h_i(q_i, z) = \beta n$. Thus q'' is the contribution that i would most prefer that everyone should make. (That is, if I is the group containing everyone, $q_i^I = q''$.)

Equilibrium occurs if and only if q_i, the common contribution of every individual, lies in the range $q' \leqslant q_i \leqslant q''$. If $q_i < q'$, every individual would find that self-interest dictated a larger contribution, even if he had no expectation that others would reciprocate. If $q_i > q''$, every individual would be contributing more than he was obliged to. (It is obvious that no one can be under an obligation to I – the group containing everyone – to contribute more than q''. It can be shown[1] that this entails that no one can be under such an obligation to *any* group.) But if $q' \leqslant q_i \leqslant q''$, everyone is obliged to reciprocate everyone else's contributions, while neither reciprocity nor self-interest dictates that anyone should contribute more than he actually does.

Four features of this special case are worth noting. First, equilibrium exists. Second, it is not unique.[2] Third, one equilibrium is Pareto efficient. (If everyone contributes q'', each person's MRS between the public good and effort is equal to βn. Equivalently, the sum of individuals' MRS's between effort and the public good is equal to the marginal rate of transformation between effort and the public good. This equality corresponds with Samuelson's (1954) well-known efficiency condition.) Fourth, every other equilibrium involves under-supply of the public good. These inefficient equilibrium states are ones in which everyone would contribute more if only he knew that the others would too, but in which no one will make the first move. They are instances of the 'assurance problem' (Sen, 1967) as opposed to the n-person prisoner's dilemma problem (in which no one would contribute more *even if* he knew that everyone else would).

V. EQUILIBRIUM IN THE GENERAL CASE

I shall now consider the general case in which preferences may differ between people, and investigate how far the conclusions of Section IV can be generalised.

First I shall prove the following result:

Result 1. Given any admissible production function and any admissible pattern of preferences, an equilibrium vector of contributions exists.

[1] See Result 6, proved in the appendix.

[2] There is one exception: equilibrium is unique if $q'' = 0$. This is the case in which it is Pareto efficient not to produce the public good at all.

I shall prove this by describing a theoretical procedure for generating an equilibrium vector of contributions, given any pattern of individual's preferences and any production function.

Begin with the vector of contributions $\mathbf{q} = (0, \ldots, 0)$. Consider a small increment of effort, δq. Let $\mathbf{q}'$ be the vector of contributions in which every member of some group G' contributes δq and in which everyone else contributes nothing; and let G' be the largest[1] group that has the property that every member of the group prefers $\mathbf{q}'$ to $\mathbf{q}$. In other words, G' is the largest group that can be formed in which every member has an obligation to contribute at least δq, provided that all the other members do the same. Now consider a further increment of effort. Let $\mathbf{q}''$ be the vector of contributions in which every member of some group G'' (where G'' is a subset of G') contributes δq more than he did in the vector $\mathbf{q}'$, and in which everyone else contributes as much as he did in $\mathbf{q}'$; and let G'' be the largest group such that every member prefers $\mathbf{q}''$ to $\mathbf{q}'$. This process can be repeated again and again until a vector $\mathbf{q}^*$ is reached at which it is impossible to find any group of people who would be willing to contribute more (conditional on the other members of the group doing the same). This vector $\mathbf{q}^*$ is an equilibrium.

Why must $\mathbf{q}^*$ be an equilibrium? First consider the case of someone, say i, who is not in G' and who therefore contributes nothing. Since G' is, by definition, the largest group of people who can be under an obligation to contribute δq, i cannot be under this obligation; so his contribution of nothing is compatible with his obligations. Next consider the case of someone, say j, who *is* a member of G'. We know that he is obliged to contribute at least δq, provided that everyone else in G' does the same. And the vector $\mathbf{q}^*$ has the property that everyone else in G' *does* contribute at least δq; so j is obliged to contribute at least δq. By repeating this analysis for all the groups G'', G''', ..., we can deduce that the vector $\mathbf{q}^*$ is one in which everyone contributes just as much as he is obliged to, and no more. In other words, we can deduce that $\mathbf{q}^*$ is an equilibrium.

Of course, $\mathbf{q}^*$ need not be the *only* equilibrium, since – as I showed in Section IV – equilibrium is not necessarily unique. From now on I shall investigate properties that are common to *all* equilibrium vectors of contributions.

One important property of equilibrium can be described in terms of the *cumulative contribution function* $\Phi(.)$. This is defined for any given vector of contributions by:

$$\Phi(\xi) = f[\sum_i \alpha_i \min(q_i, \xi)]. \tag{7}$$

Thus $\Phi(\xi)$ is the amount of the public good that would have been produced if everyone who actually contributes more than ξ had instead contributed ξ, everyone else's contribution remaining unchanged. It follows from this definition and from the properties of the production function that $\Phi(.)$ is continuous, non-decreasing and concave. The following result can be proved (the proof is in the appendix):

Result 2. If $\mathbf{q}$ is an equilibrium vector of contributions, then for each individual i, q_i is the value of ξ that maximises $u_i[\xi, \Phi(\xi)]$.

[1] The largest group is unambiguously defined: every other group with this property is a subset of G'.

Thus in equilibrium, i's contribution is the ξ-coordinate of the point on the $\Phi(\xi)$ function that he most prefers. It is as if he chose a contribution to maximise his utility, taking $z = \Phi(q_i)$ as a hypothetical budget constraint.

This result allows us to make some deductions about how contributions will differ between individuals in any equilibrium. Given any cumulative contribution function $\Phi(.)$, how much each individual contributes depends on the nature of his preferences in (q_i, z) space. Roughly speaking, those with the strongest preferences for the public good relative to effort will tend to make the largest contributions. More formally:

Result 3. Let i and j be any two individuals such that for all ξ and z, $h_i(\xi, z) < h_j(\xi, z)$. Then in any equilibrium $q_i \geqslant q_j$.

This result follows straightforwardly from the concavity of $\Phi(.)$. One implication of this is that if effort is measured in *absolute* money units, richer individuals will tend to contribute more than poorer ones. (If i and j have identical preferences defined over combinations of *private consumption* and the public good, and if i has the higher income, then $h_i(\xi, z) < h_j(\xi, z)$ will be true for all ξ and z.)

It is difficult to say much more than this without making more specific assumptions about individuals' preferences. However, there is reason to expect the frequency distribution of contributions to be rather skewed.[1] Each individual chooses his contribution as though his budget constraint was $z = \Phi(q_i)$; and $\Phi(.)$ is a concave function. So it is as though the marginal price of the public good, measured in units of effort, increases as a person's contribution increased. This effect is particularly marked towards the top end of the frequency distribution. Thus this distribution is unlikely to have much of an upper tail.

It is also possible to make some deductions about how one person's contribution is influenced by other people's. Consider any individual i, and take the contributions of all other individuals as given. If i acts on the reciprocity principle, how will his contribution change as other people's contributions change? The answer is as follows:

Result 4. Suppose that person j increases his contribution from q_j' to q_j'', while the contributions of all other persons apart from i remain constant. Let q_i' and q_i'' be the contributions that i would make in the two cases. Then (i) $q_j'' < q_i'$ entails $q_i'' \leqslant q_i'$; (ii) $q_j' > q_i'$ entails $q_i'' = q_i'$; and (iii) $q_j' = q_i'$ entails $q_i'' \geqslant q_i'$.

(This is proved in the appendix.) Case (iii) is the one in which j's extra contribution can create an additional obligation for i; in this case, an increase in j's contribution can lead to an increase in i's.

Finally, consider the efficiency properties of equilibrium. An equilibrium is characterised by *under-supply* if it is possible to generate a Pareto improvement by increasing the production of the public good, and by *over-supply* if it is possible to generate such an improvement by decreasing production. The following result can be proved (the proof is in the appendix):

[1] This idea was first suggested to me by Howard Margolis.

Result 5. Every equilibrium vector of contributions $\mathbf{q}$ is either Pareto efficient or is characterised by under-supply; and Pareto efficiency occurs if and only if $\mathbf{q} = (q', \dots, q')$, where $q' = q_i^I$ for all i, I being the set of all persons.

In other words, Pareto efficiency is possible only if, were everyone to be asked to choose a single contribution for everyone in the community, they would all opt for the same contribution. (Pareto efficiency is still only *possible* and not *certain* because of the assurance problem.) In every other case, equilibrium is a state of under-supply.

VI. RECIPROCITY AND OTHER THEORIES OF NON-SELFISH BEHAVIOUR

So far I have presented the theory of reciprocity in a highly abstract form. It is now time to consider, in a more concrete way, what sort of behaviour the theory predicts, and how these predictions differ from those of other theories of non-selfish behaviour.

(i) *The mixed success of voluntary activity*

Even for a society of identical individuals, the theory of reciprocity does not predict that the free-rider problem *will* be solved. Because of the assurance problem, a society of moral citizens can get locked into an equilibrium in which no one contributes anything towards a public good – even though everyone would prefer that everyone contributed. The theory says only that the free-rider problem *can* be solved.

This may seem to be a weakness of the theory, but we have to recognise that although voluntary activity sometimes succeeds in supplying public goods in significant amounts, it often fails. A fully satisfactory theory ought to be able to explain both observations. My theory is at least *consistent* with both of them. Since the nature of the assurance problem is clear enough, it might be possible to extend the theory to explain the sort of initiative that overcomes it.

The mixed success of voluntary activity is, however, a serious problem for the theory of unconditional commitment. Anyone who acted in accordance with that theory would contribute towards the supply of every public good from which she benefitted; so the theory cannot account for the observation that some public goods are supplied through voluntary activity and some are not. Margolis's theory of altruism does not have this problem, but only because it does not try to explain the content of an individual's *G*-preferences. In particular, Margolis does not require any particular correspondence between *S*-preferences and *G*-preferences (Margolis, 1982, pp. 48 and 98–101).

(ii) *Which activities does an individual support?*

Reciprocity theory – in common with the theory of unconditional commitment – makes a very specific prediction about *which* public-good producing activities an individual will support. According to either of these theories, an individual will contribute towards the production of those public goods from which she derives benefits, and not – unless for reasons unconnected with the theory – towards the production of other public goods. (Here I interpret 'benefit' in terms of self-interest, so this is a genuine prediction and not just a tautology.) Once we recognise that self-interest does not usually require *any* voluntary contributions towards public goods (free-riding being the best strategy for a selfish individual) it is by no means self-evident that individuals will support only those activities from which they derive benefits.

Margolis's theory would certainly not predict this. If a person is choosing *which* public goods to contribute towards, her *S*-preferences are irrelevant. Since it is in her self-interest to take a free ride, her *S*-self will not devote *any* resources to the production of public goods; any contributions she makes must come from the resources allocated to her *G*-self. But her *G*-self takes a disinterested view of the welfare of society, and there is no reason for it to favour an activity merely because she happens to benefit from it. The logic of the theory requires that she should allocate her contributions between public goods and other 'group-oriented' activities so as to maximise social welfare (as she conceives it). If the values of the resources available to her *G*-self is small relative to the extent of a typical public-good producing activity, *G*-utility is likely to be maximised by plumping for a single activity – the one she believes generates the largest increment of social welfare from a marginal contribution. For most people, I suggest, this single activity would be a humanitarian charity. It is hard to see how mundane public goods, like the union that represents workers at the firm where you happen to work or the playgroup to which you happen to send your children, could take priority over, say, famine relief in the third world or support for families with mentally handicapped children. Indeed, it is not even clear why that particular union and that particular playgroup should take priority over other unions and other playgroups.

So if there is any marked tendency for people to contribute to the particular public goods from which *they* derive benefits, this is evidence in favour of a theory of cooperation and against a theory of altruism.

(iii) *The under-supply of public goods*

Reciprocity theory predicts that voluntary activity will lead to the under-supply of public goods – 'under-supply' being interpreted in the Paretian sense. This is a clear-cut prediction which has been confirmed in at least one experimental test (Marwell and Ames, 1979, 1980).

It is tempting to say that this prediction is an obvious one; but it is obvious only if we assume that people always free-ride when they can, and we know that that assumption is false. Neither of the other two theories of non-selfish behaviour predicts under-supply. According to the theory of unconditional commitment, each person contributes the amount she would want everyone to contribute, and there is no obvious reason why this should bias the supply of public goods in the direction of under-supply. Since Margolis's *G*-preferences are independent of *S*-preferences, his theory has nothing to say on this question.

(iv) *Homogeneous and heterogeneous communities*

According to reciprocity theory, Pareto efficiency is possible in a community of identical individuals; but if individuals are allowed to differ, efficiency is possible only as the result of a most unlikely coincidence (see Result 5). This suggests the following conjecture: the more homogeneous a community is in respect of incomes and tastes, the more closely it can approach Pareto efficiency, and the greater will be its success in producing public goods through voluntary activity. Roughly speaking, the problem for the heterogeneous community is that those people with the strongest preferences for the public good will not contribute as much as *they* would like everyone to contribute – because the others (quite justifiably) would not reciprocate.

With the theory of unconditional commitment this problem does not arise, because each person contributes as much as she would like everyone to contribute – irrespective of what the others do. There seems no reason, then, why heterogeneity should tend to lead to under-supply. Once again, this is an issue on which Margolis's theory makes no prediction at all.

It is often suggested that small and homogeneous communities are the most effective in inspiring group loyalty and a willingness to act according to moral principles (e.g. Buchanan, 1978). It seems that reciprocity theory may be able to reach a rather similar conclusion by an entirely different route. People in heterogeneous communities may be just as willing to meet their moral obligations to one another as people in homogeneous ones, and yet the heterogeneous communities may still be less capable of supplying public goods through voluntary cooperation.

(v) *The effect of changes in other people's contributions*

Possibly the most distinctive prediction of reciprocity theory is contained in Result 4: if person j's contribution is initially the same as person i's, an increase in j's contribution will tend to bring about an *increase* in i's. The idea that each person tends to contribute more as others contribute more has, I think, some intuitive plausibility; and there is some experimental evidence that people are induced to contribute more to public goods and charitable activities by seeing other people contributing (Bryan and Test, 1967). Such a relationship between contributions has occasionally been *assumed* in theories of voluntary activity (e.g. Cornes and Sandler, 1984), but it has not been *explained*. Indeed, most theories of non-selfish behaviour predict that no such relationship will exist.

The theory of unconditional commitment would obviously predict that each individual's contributions would be independent of everyone else's. The logic of Margolis's theory points to an inverse relationship. (This is a characteristic prediction of theories of altruism: compare Schwartz (1970) and Becker (1974).) Suppose, for example, that each of i's colleagues has contributed £5 towards the office Christmas party, and that i has done the same. If Margolis is right, i's G-self is allocating its share of resources between various group-oriented activities so as to maximise G-utility. If other people's contributions to the party increase, the marginal G-utility of expenditure on this activity must fall; so the rational response for i's G-self is to divert some of its resources from the party to other activities.

(vi) *The shape of the frequency distribution of contributions*

If reciprocity theory is correct, the frequency distribution of contributions is likely to be skewed, with little or no upper tail. Very roughly, what this amounts to its that no one is willing to contribute much more than anyone else – which, given the logic of the reciprocity principle, is natural enough. This would seem to be another distinctive prediction of reciprocity theory.

Exactly what this prediction means in empirical terms depends on how effort is measured. For example, if effort is measured by *absolute* money contributions, we should expect to find the frequency distribution of money contributions to be skewed, with the income elasticity of contributions being close to zero at high levels of income. If instead effort is measured by *relative* money contributions, it is the distribution of 'contributions as a proportion of income' that will be skewed; the income elasticity of absolute contributions is likely to be close to unity. This uncertainty about the definition of effort makes it difficult to use existing econometric studies of charitable giving[1] to test reciprocity theory. A further problem is that these studies typically involve a great deal of aggregation, while the predictions of reciprocity theory concern contributions to a single public good.

A better way to test the theory would be to find a case where a particular good is public to a relatively small group of people who have similar incomes, but different tastes for the good in question. (The smallness of the community enables each person to know what the others are doing; the similarity of their incomes reduces the ambiguity of the concept of effort; and the diversity of taste provides a reason for expecting contributions to vary.) If the public good is produced from voluntary contributions of money, reciprocity theory predicts a skewed distribution of contributions. The other two theories do not.

VII. CONCLUSION

The economic analysis of non-selfish behaviour is still in its infancy: there is no unified theory that can explain all, or even most, of the observed regularities in such behaviour. Some apparently promising starting points have been suggested by various writers, but that is all. The theory of reciprocity is relatively simple; it rests on a principle of practical morality that seems to have a strong common-sense appeal; it is compatible with both the observed successes and the observed failures of voluntary cooperation; and it generates a wide range of testable predictions. It deserves to be taken seriously, as one among several avenues of enquiry into a particularly puzzling area of economics.

University of Newcastle upon Tyne

Date of receipt of final typescript: February 1984

[1] The main findings of several of these studies are reviewed in Sugden (1982).

Appendix: proofs of the results

As a first step, it is useful to prove an additional result:

Result 6. Take any vector of contributions $\mathbf{q}$ (not necessarily an equilibrium one). Let G be any non-empty set of persons, let H be any subset of G that contains person i, and let ξ^* be any level of contributions such that for all persons j in G (excluding i himself), $q_j \geqslant \xi^*$. Then $q_i^H \geqslant \xi^*$ entails $q_i^G \geqslant \xi^*$.

To prove this, consider the functions $F(G, \xi)$ and $F(H, \xi)$. Since $H \subset G$ and $q_j \geqslant \xi^*$ for all j in the set $G-H$, we know (a) that $F(G, \xi^*) \leqslant F(H, \xi^*)$ and (b) that $F_\xi(G, \xi^*) \geqslant F_\xi(H, \xi^*)$. Because of ($a$) and assumption (4), we know (c) that $h_i[\xi^*, F(G, \xi^*)] \leqslant h_i[\xi^*, F(H, \xi^*)]$. Suppose that $q_i^H \geqslant \xi^*$. This entails (d) that $h_i[\xi^*, F(H, \xi^*)] \leqslant F_\xi(H, \xi^*)$. Combining ($b$), ($c$) and ($d$), $h_i[\xi^*, F(G, \xi^*)] \leqslant F_\xi(G, \xi^*)$; and this entails that $q_i^G \geqslant \xi^*$.

It is also useful to adopt the following notation. Consider any vector of contributions $\mathbf{q}$. Let q^1 be the smallest non-zero contribution, let q^2 be the second smallest, and so on to q^m which is the largest. (For example, if $n = 5$ and $\mathbf{q} = (5, 0, 1, 1, 3)$, then $q^1 = 1$, $q^2 = 3$ and $q^3 = q^m = 5$.) Let $q^0 = 0$. For any $\xi \geqslant 0$, let $G(\xi)$ be the group of people whose contributions are strictly greater than ξ. Now (using the definition (6)) consider the function $F[G(q^l), \xi]$ for each $l = 0, \dots, m$. For all $l = 0, \dots, m-1$, this function is continuous, increasing and concave in ξ. If $l = m$, $G(q^l)$ is empty and for all $\xi \geqslant 0$, $F[G(q^l), \xi] = z$ where z is the amount of the public good produced from the contributions $\mathbf{q}$.

Results 1 and 3 are proved in the main text. The proofs of the other results are as follows:

A proof of Result 2

Consider any equilibrium vector $\mathbf{q}$. Consider any person i, and let his contribution be the lth smallest (i.e. $q_i = q^l$).

First suppose that $q_i > 0$ and (in contradiction to Result 2) that the value of ξ that maximises $u_i[\xi, \Phi(\xi)]$ is *less than* q_i. It follows from the definitions (6) and (7) that in the range $q^{l-1} \leqslant \xi < q^l$, $\Phi(\xi) = F[G(q^{l-1}), \xi]$. So the value of ξ that maximises $u_i\{\xi, F[G(q^{l-1}), \xi]\}$ must also be less than q_i. But this entails that i is not under an obligation to $G(q^{l-1})$ to contribute as much as q_i. Because of Result 6, he can have no such obligation to *any* group; the original supposition generates a contradiction.

Now suppose that the value of ξ that maximises $u_i[\xi, \Phi(\xi)]$ is *greater than* q_i. In the range $q^l \leqslant \xi < q^{l+1}$, $\Phi(\xi) = F[G(q^l), \xi]$. So the value of ξ that maximises $u_i\{\xi, F[G(q^l), \xi]\}$ must also be greater than q_i. But this entails that i is under an obligation to the group that contains himself and the members of $G(q^l)$ to contribute more than q_i: the second supposition has generated a contradiction too. Thus the value of ξ that maximises $u_i[\xi, \Phi(\xi)]$ must be exactly q_i.

A proof of Result 4

Given the initial vector of contributions $\mathbf{q}'$, consider the functions $F(H, \xi)$ and $F(J, \xi)$ where H is the set of all people who contribute q_i' or more and J is the set containing i and all people who contribute more than q'_i. We know that i is obliged to contribute q'_i but no more; this entails $q_i^H \geqslant q'_i$ and $q_i^J \leqslant q'_i$. Now consider how the functions $F(H, \xi)$ and $F(J, \xi)$ will shift if person j increases his contribution from q'_j to q''_j. In case (i), $q''_j < q'_i$; so j is never a member of H or J. The increase in j's contribution shifts both of the functions upwards while reducing (or at least, not increasing) their slopes at each value of ξ. (The reduction in their slopes is a result of the concavity of the production function.) Thus (because of (4)) the values of q_i^H and q_i^J must both fall. This entails $q''_i \leqslant q'_i$. In case (ii), $q'_j > q'_i$ so j is always a member of both H and J. The increase in contribution has no effect on either of the functions, and so i's contribution does not change. In case (iii), $q'_j = q'_i$, so before the increase in j's contribution he is a member of H but not J; after the increase, he is a member of *H and J*. The increase has no effect on $F(H, \xi)$, and so i's obligation to contribute at least q'_i is unaffected. But the increase also causes $F(J, \xi)$ to shift. At $\xi = q'_i$, the value of this latter function stays constant but its gradient increases; so it is possible that the new value of q_i^J is greater than q'_i. Thus i may now be obliged to contribute more than q'_i.

A proof of Result 5

Consider any equilibrium vector of contributions $\mathbf{q}$. Let r be defined by $r = \sum_i \alpha_i q_i$ and let z be the amount of the public good produced from $\mathbf{q}$; thus $z = f(r)$.

For any person i, an extra contribution of $1/h_i(q_i, z)$ would exactly offset the extra utility he would derive from a marginal unit of the public good. If each person i increased his contribution by $1/h_i(q_i, z)$, the supply of the public good would increase by $f'(r) \sum_i [\alpha_i/h_i(q_i, z)]$. If this increase in supply is exactly one unit, $\mathbf{q}$ is Pareto efficient. If the increase is more than one unit, there is under-supply; if it is less than one unit, there is over-supply (unless $z = 0$, in which case there is Pareto-efficiency).

Let J be the set of people who contribute q^m; thus $J = G(q^{m-1})$. Consider any person j in J. Using Result 6, j must be under an obligation to J to contribute q^m. Thus the value of ξ that maximises $u_j[\xi, F(J, \xi)]$ must be no less than q^m. This entails that

$$h_j(q_j, z) \leqslant f'(r) \sum_{j \in J} \alpha_j. \tag{A 1}$$

(The right-hand side of the inequality (A 1) is the slope of the function $F(J, \xi)$ evaluated at $\xi = q^m$.) Inverting both sides of (A 1) and then multiplying by α_j,

$$\alpha_j/h_j(q_j, z) \geqslant \alpha_j/f'(r) \sum_{j \in J} \alpha_j. \tag{A 2}$$

Summing (A 2) over all persons in J and re-arranging:

$$f'(r) \sum_{j \in J} [\alpha_j/h_j(q_j, z)] \geqslant 1. \tag{A 3}$$

Notice that (A 3) holds as a strict equality if and only if $q_j^J = q^m$ for all persons j in J.

Let I be the set of all individuals. Since $\alpha_i/h_i(q_i, z) > 0$ for all persons i, it follows from (A 3) that

$$f'(r) \sum_{i \in I} [\alpha_i/h_i(q_i, z)] \geqslant 1. \tag{A 4}$$

(A 4) holds as a strict equality if and only if $J = I$ and $q_i^I = q^m$ for all persons i. This amounts to a proof of Result 5.

References

Arrow, K. J. (1981). 'Optimal and voluntary income distribution'. In *Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Bergson* (ed. S. Rosefielde). Cambridge University Press.

Becker, G. S. (1974). 'A theory of social interactions.' *Journal of Political Economy*, vol. 82, pp. 1063–93.

Bryan, J. H. and Test, M. A. (1967). 'Models and helping: naturalistic studies in aiding behaviour.' *Journal of Personality and Social Psychology*, vol. 6, pp. 400–7.

Buchanan, J. M. (1978). 'Markets, states and the extent of morals.' *American Economic Review*, vol. 68, Papers and Proceedings, pp. 364–8.

Collard, D. A. (1978). *Altruism and Economy*. Oxford: Martin Robertson.

—— (1983). 'Economics of philanthropy: a comment.' Economic Journal, vol. 93, pp. 637–8.

Cornes, R. and Sandler, T. (1984). 'The theory of public goods: non-Nash behaviour.' *Journal of Public Economics*, vol. 23 (forthcoming).

Guttman, J. M. (1978). 'Understanding collective action: matching behaviour.' *American Economic Review*, vol. 68, Papers and Proceedings, pp. 251–5.

Harsanyi, J. C. (1980). 'Rule utilitarianism, rights, obligations and the theory of rational behaviour.' *Theory and Decision*, vol. 12, pp. 115–33.

Laffont, J.-J. (1975). 'Macroeconomic constraints, economic efficiency and ethics: an introduction to Kantian economics.' *Economica*, vol. 42, pp. 430–7.

Margolis, H. (1982). *Selfishness, Altruism and Rationality*. Cambridge University Press.

Marwell, G. and Ames, R. (1979). 'Experiments on the provision of public goods. I. resources, interest, group size, and the free rider problem.' *American Journal of Sociology*, vol. 84, pp. 1335–60.

—— (1980). 'Experiments on the provision of public goods. II. provision points, stakes, experience, and the free rider problem.' *American Journal of Sociology*, vol. 85, pp. 926–37.

Samuelson, P. A. (1954). 'The pure theory of public expenditure.' *Review of Economics and Statistics*, vol. 36, pp. 387–9.

Schwartz, R. A. (1970). 'Personal philanthropic contributions.' *Journal of Political Economy*, vol. 78, pp. 1264–91.

Sen, A. K. (1967). 'Isolation, assurance and the social rate of discount.' *Quarterly Journal of Economics*, vol. 81, pp. 112–24.

Sugden, R. (1982). 'On the economics of philanthropy.' Economic Journal, vol. 92, pp. 341–50.