The Limits to Altruism - A Survey

Richard Povey, Hertford College, Oxford University

November 10, 2014

> There is much confusion of the ideal that a person ought to be allowed to pursue his own aims with the belief that, if left free, he will or ought to pursue solely his selfish aims.
>
> (Hayek, 1960)

> In my view the ideal society would be one in which each citizen developed a real split personality, acting selfishly in the market place and altruistically at the ballot box.
>
> (Meade, 1973)

1.1 *Overview*

This survey is concerned with altruism, and its relationship to economic theory. The key tasks are to summarise and marshal the evidence that altruism exists as a significant empirical phenomenon, and to make a contribution to the normative analysis of altruism by bringing together a number of theoretical strands from economics and other social sciences. The central questions to be considered are why particular patterns of altruistic motivation and behaviour exist, and which conceivable patterns would be most beneficial. The answers to these two questions are closely connected if there is a theoretical presumption that cultural norms emerge at least partly due to their social functionality. We clearly live in a world of imperfect altruism. The central question is *why* this should be the case. Are individuals imperfectly altruistic because we live in an imperfect world, or could it be that these imperfections are somehow functional for society?

The appropriate role of human altruism in social scientific theory and methodology is subtle and controversial. Firstly, altruism can be defined in numerous incompatible ways, particularly when we

compare the approaches of economists, anthropologists, sociologists, psychologists and biologists. A central factor which commonly causes confusion here is the distinction between altruism as a concept for classifying observed behaviour, and altruism as a theory of motivation. There are also important distinctions between economic and psychological theories of altruistic motivation, due to their differing criteria for what constitutes an explanation for a particular phenomenon.

A second key area of debate is the normative analysis of altruism. The temptation that some approaches have fallen into is to assume that altruism on the part of individuals is always good for the group, and thus that minimising individual selfishness is the most important role for the social structure. Economics has, at least since Adam Smith, had a fairly clear understanding that a key litmus test for the effectiveness of a social and economic system is its ability to correct for, and harness, the limitations upon individual altruism (Smith, 1976). This is just as important as ensuring that individual selfishness is kept constrained within certain boundaries. For example, the effective rule of law is required for a well-functioning market economy, and this requires that individuals refrain from harming others indiscriminately whenever they can gain selfishly by doing so. In other words, selfishness must, in order to act as a force for social good, be constrained to operate within a broader moral framework. However, the relationship between the positive assumption of rational self-interest as a parsimonious explanatory framework, and the normative argument that self-interest is a force that can be harnessed for the common good, has remained inadequate and blurred.

1.2 *Altruism: A Road Map*

If we wish to achieve clarity in an exploration of the past and future potential role of altruism in economic theory specifically, we must have a clear conception of the distinctions and connections between a normative concept of social efficiency (the common good), models of individual altruistic motivation (which are explanatory rather than normative), and a way of classifying the degree of altruism exhibited in various behaviours based upon observable criteria. The fact that apparently altruistic behaviour does not necessarily imply altruistic motivation and that altruistic motivation does not necessarily imply social efficiency suggests eight logical possibilities for the interpretation of a particular behaviour:

(1). Apparently altruistic behaviour which is altruistically motivated and is socially beneficial.

(2). Apparently altruistic behaviour which is not altruistically motivated and is socially beneficial.

(3). Apparently non-altruistic behaviour which is altruistically motivated and is socially beneficial.

(4). Apparently non-altruistic behaviour which is not altruistically motivated and is socially beneficial.

(5). Apparently altruistic behaviour which is altruistically motivated and is socially detrimental.

(6). Apparently altruistic behaviour which is not altruistically motivated and is socially detrimental.

(7). Apparently non-altruistic behaviour which is altruistically motivated and is socially detrimental.

(8). Apparently non-altruistic behaviour which is not altruistically motivated and is socially detrimental.

This system of categorization is useful in maintaining a separation between the behavioural, motivational and normative dimensions to altruistic phenomena. The different cases are illustrated graphically in figure 1.1. The models examined later in the survey can be fitted into it. The way in which the categories are applied, however, will vary depending on one's theoretical perspective.

Welfare economic theory already has a well-defined set of tools for judging when changes in behaviour are socially beneficial and socially detrimental, provided one is philosophically willing to make inter-personal comparisons of utility from consuming different goods (Sen, 1974) (d'Aspremont & Gevers, 1977) (Gevers, 1979) (Roberts, 1980). The concepts of a behaviour being "apparently altruistic" and "altruistically motivated" are, however, more problematic. In asking whether a behaviour appears to be altruistic, we must ask who or what appears to be paying the cost, and who is benefiting. Dawkins' selfish gene theory (Dawkins, 1976), for example, seeks to categorise all cases of apparent altruism by individual organisms as the manifestation of the entirely selfish adaptive "behaviour" of genes. In this case, there is strictly speaking no apparent or actual altruism at the gene level. So, if we are taking the "gene's eye view", then cases (1), (2), (3), (5), (6) and (7) drop out. The only question that remains to be answered in the analysis of a particular behavioural phenotype is therefore whether the interaction of competition of selfish genes leads to a "socially beneficial" or "socially detrimental" outcome.

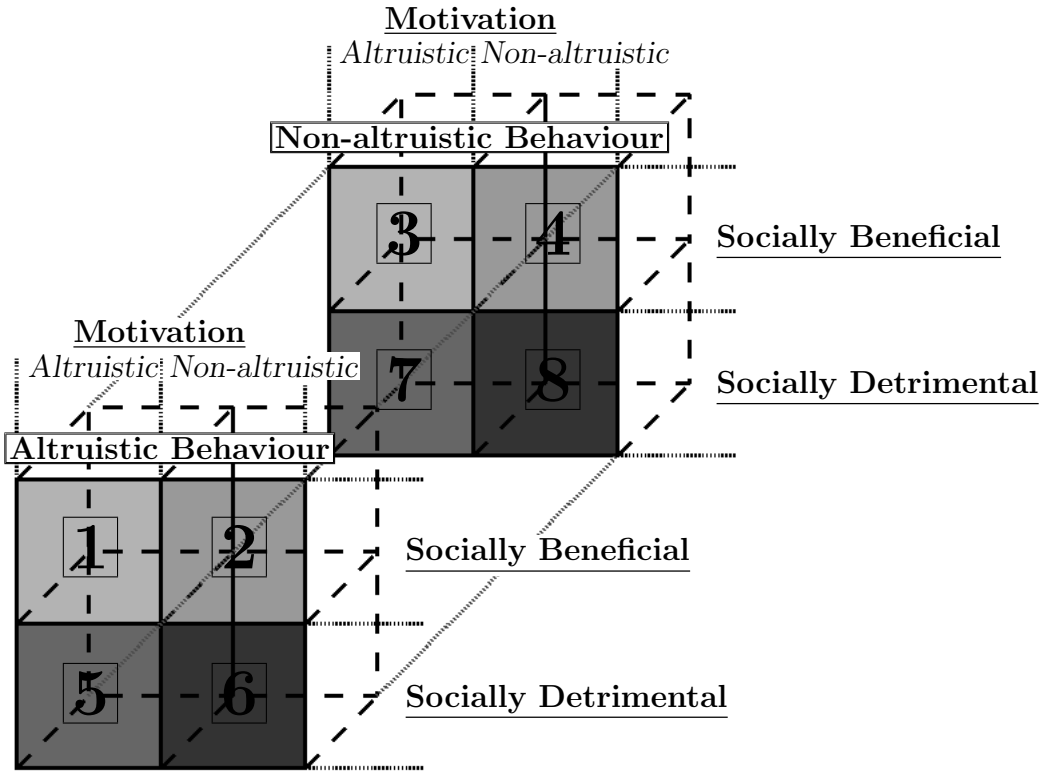

Figure 1.1: Road-map: cases of altruism

To give just one example from evolutionary biology, when animals of the same species fight for territory and status they do not usually fight to the death. In fact, fights usually appear to the human eye as something of a spectacle rather than a cut-throat struggle for survival. It was fashionable among biologists in the earlier part of the twentieth century to interpret this as a social convention which evolved for the good of group stability. However, evolutionary game theory shows that the evolution of such "conventions" can be more cleanly explained as the result of the strategic interaction of selfish animal organisms which are entirely concerned with their own survival. In a duel, the weaker animal will usually give up long before the viciousness of the conflict escalates very much, once it becomes reasonably clear what the outcome will be. It does this in its own self-interest, since there is no point "throwing good money after bad" once it has enough evidence about the strength of its opponent to "know" that it will lose. (Obviously, we are not necessarily talking about a conscious process but about evolved "rules of thumb" hardwired into the animal brain.) Nonetheless, it remains the case that these animal "social conventions" do contribute to the stability of hierarchical animal groups, and therefore indirectly to the safety of all of their members.

Gene selfishness is a distinct concept from that of the biological selfishness of the individual organism. Looked at from this perspective, there is a clear definition that an altruistic action increases the likelihood of other organisms to reproduce at the expense of the altruistic individual. This can be compatible with the selfish gene theory, because it can be brought about via the selfish genes seeking to aid copies of themselves in other animal bodies by damaging the narrower biological interests of the particular body they currently inhabit. Note also that this biological definition is distinct from the modelling of the psychological processes which lead to a particular behaviour. Evolutionarily altruistic acts committed in order to avoid a feeling of guilt, for instance, are psychologically egoistic.

None of the above definitions of apparent altruism are adequate for the social sciences in general, or economics in particular, because the connection between behaviours which economically benefit others and fitness in the biological evolutionary sense is no longer present in contemporary societies. Economically poor people on average breed more and thus have a higher evolutionary fitness in narrowly biological terms (Simon, 1993). To give another example, economists would wish to model charitable giving as altruistic, even though it may be performed for psychologically egoistic reasons.

A theoretical framework for distinguishing between economically altruistic and non-altruistic acts must specify a number of different goods which enter into each individual's utility function and distinguish those which are directly beneficial to the individual's economic well-being. This part of the individual's utility function is usually referred to as **felicity**. This is not the same as the biologist's definition of the fitness of either individual organisms or genes, because economic well-being occurs within a human culture that evolves with a degree of autonomy from the biological sphere. Cultural evolution, popularised in the idea of memes (Blackmore, 1999), is made possible because of the level of complexity to which biological evolution has developed the human body, and particularly the brain (Boyd & Richerson, 1982) (Soltis et al., 1995). However, memetic evolution occurs according to its own set of rules. Memes such as suicide bombing can spread culturally even though they reduce the biological fitness of the individuals and groups who adopt them.

The individual in economic theory is a socially constructed cultural and theoretical entity, and thus distinct from the individual organism of biology. This is sometimes obvious such as when economists treat firms or families as if they were rational individuals. However, even if the fundamental explanatory unit in a particular theory is the modelling of individual people's decisions, there is still not a simple

overlap between "economic man" and "biological man". To use the terminology of Thomas Kuhn (Kuhn, 1970), the concepts used in the two disciplines are incommensurable. Both biologists and economists talk of "individuals" but the meaning of these concepts depends holistically upon the web of other concepts in the theoretical scheme, such as rationality or utility. To return to our practical concern over which goods to consider "economically fundamental", the ability of an individual to prosper in a market society depends upon more goods than those which determine narrow biological fitness. Therefore, although food and shelter should clearly enter into an individual's economic felicity, so also should biologically non-beneficial goods such as education.

Not all goods which an economically altruistic individual gains utility from can be counted in the felicity function, however. For example, if we define "lack of guilty feeling" as a good entering the individual's felicity function, then we will define away acts that we would wish to consider economically altruistic. Therefore, we instead model individuals as maximising weighted additive combinations of their own felicity and the felicity of other individuals. We are not concerned with the psychological motivations that lead individuals to behave in this way.[1]

An appealing analogy may exist between the distinction between economically fundamental goods and those which are not included in individuals' felicity functions and the Marxist concept of the economic base and the superstructure.[2] Ultimately the relative success of individuals is determined by their consumption of economically fundamental resources. However, the production and distribution of these resources within a particular group is shaped by individual preferences which can include a concern for the felicity of other individuals. Whether the psychological mechanism for this is guilt, concern for others, duty to God or whatever, these can be modelled as economically altruistic preferences. The economic base is served to a greater or lesser extent by particular types of preferences within the moral superstructure. The success of these morals in helping individuals and groups to "get on" determines which will spread and propagate. Here, the Marxist view of social change has been woefully inadequate, in that it fails to take proper account of the perennial tension between the individual and the group. Modern cultural evolution theory, however, offers a way to overcome this limitation. Ultimately, the "base" shapes the "superstructure". However, the superstructure influences the evolution of the base and has a degree of autonomy from it at any one time (Cohen, 1978).

[1] Henceforth, this particular concept of altruism shall be referred to simply as "altruism", but the caveats expressed so far about the differences compared to other definitions should be kept in mind.

[2] Note that this parallel with Marxist theory does not imply a commitment to Marxist political conclusions.

1.2.1 Classifications

We will now consider examples which fit into the different cases enumerated earlier. These range from true altruism, where there is no ambiguity about the social desirability or divergence between appearance and motivation of an altruistic phenomenon, to true selfishness. In between, however, are many cases where the water is muddier.

(1) Apparently altruistic behaviour which is altruistically motivated and is socially beneficial

The classic case of an altruistic act is one which benefits others at a cost to the altruistic individual, and where the total benefit conferred is greater than the cost. An individual with truly altruistic preferences would carry out such an action because it would increase their utility (since this includes the felicity of other individuals).

(2) Apparently altruistic behaviour which is not altruistically motivated and is socially beneficial

This is the case of enlightened self-interest; individuals may commit an act in a situation which at first glance appears identical to the pure altruism case. However, the act may in fact be in the individual's rational self-interest once the environment in which the decision is made is taken into account. Peter Hammond has discussed the subtleties of distinguishing between enlightened self-interest and true altruism in a game theoretic context, and suggests that the two may not be as empirically distinct as it first appears, when a refinement to Nash equilibrium is considered which enables efficient outcomes to be identified as unique non-cooperative equilibria (Hammond, 1975).

(3) Apparently non-altruistic behaviour which is altruistically motivated and is socially beneficial.

This is not a situation which is normally explored by economists, but it is important to realise, as argued by Hammond (Hammond, 1975), that observed behaviour will often be explicable both in terms of rational self-interest and rational altruism. The evidence under-determines the theory. The argument is usually made that self-interest is more parsimonious, and that therefore altruism should not be introduced unless it adds explanatory power. However, this indeterminacy can "cut both ways", and one of the central arguments of this survey will be that the abstract modelling of bone fide altruism should occupy a more comprehensive and secure position within economic theory.

(4) Apparently non-altruistic behaviour which is not altruistically motivated and is socially beneficial.

Here we have the classic case of some kind of "invisible hand" mechanism leading rational self-interested individuals to act in a socially beneficial manner. The most famous original proponents of this idea were Bernard Mandeville (Mandeville, 1732) and Adam Smith (Smith, 1976). Neoclassical economics has enabled this concept to be fleshed out and formalised in the framework of general equilibrium theory (Arrow & Debreu, 1954).

(5) Apparently altruistic behaviour which is altruistically motivated and is socially detrimental.

This is an important case which will be fleshed out in this survey, based on new research into the consequences of altruistic preferences using an abstract model of sequential punishment. More genuinely altruistic behaviour does not necessarily lead to a more socially efficient outcome. In fact, too high a level of altruism will be socially detrimental for generally applicable reasons.

(6) Apparently altruistic behaviour which is not altruistically motivated and is socially detrimental.

Perhaps the most common example of this category of phenomenon is imperfect altruism, such as localised altruism within a nation, trade union or business lobby, enabling them to overcome a collective action problem and increase group welfare at the expense of broader social efficiency. This can appear altruistic if ones' theoretical perspective fails to take into account all of the costs imposed.

(7) Apparently non-altruistic behaviour which is altruistically motivated and is socially detrimental.

This is another important case which will be explored in this survey. Altruistic motivations may lead to social outcomes which are indistinguishable from those generated in a less altruistic world, and therefore no more socially efficient.

(8) Apparently non-altruistic behaviour which is not altruistically motivated and is socially detrimental.

Into this category would fit standard cases of market failure where rational self-interested individuals are, for whatever reason, not led to behave socially efficiently by the economic environment in which they make their decisions.

1.3 *Altruism and Social Structure*

Section 1.10 surveys the empirical research which has been carried out regarding altruistic motivation in order to establish a series of "stylised facts". These can be summarised as: (A) Altruism towards other individuals is highly prevalent as a motivation for human social behaviour. (B) Real world altruism suffers from **imperfections**; meaning that real world individuals do not appear to act in a way which is consistent with putting an equal weight upon the welfare of all individuals.

Before we examine the empirical research, it is helpful to begin with some anecdotal evidence for the nature of altruism described above. Statement (A) will be denied by some, and it is undoubtedly true that methodological individualism as an approach in the social and biological sciences has greatly enhanced our understanding of Adam Smith's proposition that behaviour which serves the common good can emerge from the interactions of entirely selfish individuals. It is therefore illegitimate to assume that any example of socially functional individual behaviour must be motivated by altruism. Nonetheless, modern industrial market societies exhibit an enormous level of complex functional integration. Although economists' models of the market show why we need only enforce certain abstract rules (e.g. property rights, prevention of fraud, minimum welfare safety net) in order to achieve an efficient social order, they do not explain how the set of rules itself can be created, sustained and adapted to new situations. In other words, the provision of the legal and moral framework for a market economy is itself a public good, and subject to the free rider problem (Heckathorn, 1990).

The analysis of the repeated prisoners' dilemma has shown us that although co-operation can emerge and be sustained in small groups, it is much more difficult with a large number of individuals playing the game because of the problems of monitoring and the less credible threat of Nash-reversion or other strategies to punish the individual defector. This is why competition works, but it is also why some externalities cannot be negotiated away via Coasian bargaining. It is commonly accepted that public goods games involving a large number of interacting individuals (the global reduction of carbon dioxide emissions being a perfect case) can only be solved by compulsion or altruism. The moral and legal framework which is a prerequisite for a functioning market would seem to be a similar case. It is in each of our self-interest that everyone else respect property rights and refrain from fraud, but we each can break the rules at great gain to ourselves. Reputation formation may be able to explain some of the spontaneous formation of a market order, but the problems of underdeveloped

countries developing proper functioning market economies surely underlines the fact that some kind of effectively-run co-ordinated unit with power of compulsion (i.e. a benevolent state) is vital.

If we accept theoretically that public goods exist which can only be provided by compulsion or altruism, and that benevolently-run compulsive institutions fall into that category, the attempt to use only compulsion to explain their provision leads to an infinite regress. This is because the provision of the compulsive institution is itself a public goods problem, which can either be solved by compulsion or altruism. Sooner or later, we must postulate altruistic motivation on the part of some individuals. This does not of course settle how far back the regress goes. For example, it could be that there is only one single powerful altruistic individual in the world, who is manipulating the environment of all the others, who are entirely selfish. This, however, is not very plausible, and it seems more likely that individuals differ in their degree of altruism, with those more altruistic ones contributing to sustaining a social structure which can harness the limited altruism of the less altruistic ones (the market economy being an example of an institutional structure which achieves this).

We now come to statement (B). From the above discussion, the fact that society requires private markets, a legal system and prisons is itself powerful anecdotal evidence that individuals have greater or lesser imperfections in their altruistic perceptions and motivations. If this were not the case, we could have a society more along the lines of a democratic anarchistic society, where all individuals, due to their perfect altruism, were able to discuss, agree upon the common good and then act in concert. (Some people might see this as a real political possibility, but the key point is to be clear about the rationale concerning the type and degree of altruism that would be required of the individuals in society in order for such alternative political structures to operate efficiently.)

The institutions of a liberal democratic market society, such as markets, legislatures (and the corresponding rules to facilitate organised discussion), legal systems and law enforcement agencies can all be seen as **altruism amplification devices**; they attempt to get the best possible social outcome from the imperfect altruists who form the "human material" of the polity (Sober & Wilson, 1999). In "The Socially Optimal Level of Altruism" (Chapter 2), a simplified abstract model of this process is presented, and it is shown that altruistic preferences are ambiguous in their effect on social welfare, once they are permitted to interact with a system of social incentives, implemented via schemes of punishment.

1.4 *The Self-Interest Assumption*

The assumption of rational self-interested behaviour has been a highly fruitful one in economic theory. It has a simplicity and elegance that has allowed economics to live up to its founders' project of showing how a socially desirable spontaneous order can arise in market society which encompasses the knowledge and aims of many individuals (Hayek, 1960). Despite this, in recent years there has been a growth in the desire to move beyond the self-interest assumption. As we will see in section 1.10, some of this has taken the form of empirical work demonstrating convincingly that many specific cases of individual behaviour can be explained more adequately once the self-interest assumption is dropped.

What remains unsatisfactory is the degree to which the standard paradigm of welfare economic analysis continues to assume that whilst social institutions may ideally be designed by a benevolent utilitarian social planner, the ordinary individuals who act within them usually remain highly rational but at the same time highly selfish. The issue of whether this approach is adequate for normative analysis in economics takes us right to the heart of key issues in the philosophy of the social sciences.

Much of the initial attraction and success of the self-interest assumption, as well as the apparent explicit basis for the more recent work which has questioned or modified it, is based on a narrowly "positivistic" view of economics. This label is intended not to refer to a specific detailed position on the ontology and epistemology of social sciences, but rather to the view, most famously espoused by Karl Popper, that the value of the social sciences over and above mere superstition is their ability to make testable predictions (Popper, 1959).

In this light, the attractiveness of the self-interest assumption is clear; it provides a good basis point for a research programme in the social sciences. This concept was introduced by one of Popper's followers Lakatos. He distinguished (Lakatos, 1970) between the core of a research programme, a set of firm theoretical principles which are not questioned, and the peripheral ideas that are brought into a theory as it is tested against reality and modified. As time goes on and the possibilities of adaptation of the core principles become exhausted, it will become clear which areas of reality are difficult to explain by a particular research programme. Eventually, "progressive" research programmes become "degenerative" ones, with the peripheral assumptions becoming increasingly incoherent. With this approach, Lakatos sought to explain the discontinuities in scientific progress which Popper's theory found more difficult to explain.

The core of the standard research programme in economics could therefore be thought of as individual rationality and self interest, whereas asymmetric information or different equilibrium concepts would be peripheral components introduced and altered more freely in order to improve the fit between theory and reality. The recent increase in interest in dropping the self-interest assumption in empirical applications could be seen as a sign that this research programme has begun to exhaust the possibilities, so that in certain empirical areas it is time to consider a different research programme.

There are problems, however, with making this the philosophical justification for taking economics into the arena of altruistic motivation. Firstly, as a predictive framework, the self-interest framework is still highly productive. It would seem that if altruism is to be introduced into the body of the economic theory on this pretext, it will remain fairly ad hoc, in the sense that it will only be used when the self-interest assumption evidently fails empirically. Secondly, one senses that the attraction of altruism is not merely that it provides an alternative predictive research programme. It can be argued that economics has neglected a large area of human potential, and left the consideration of human morality and socialization largely in the hands of sociologists, when there is nothing inherent in the economist's arsenal of rigorous tools and techniques which prevents them from being applied in this area. This kind of thinking, however, requires a more subtle view of the role and potential of the social sciences in the body of human intellectual endeavour.

Kuhn took a more radical view of progress in the natural sciences than Popper and Lakatos. He argued (Kuhn, 1977) that the decision between what Lakatos would have called "progressive" and "degenerative" research programmes (the process which Kuhn referred to as "paradigm shift") could never be nailed down to evidence in such a simple way that all scientists could agree which programme provided the best potential for future development. Whenever paradigm shifts occur, there are competing considerations or scientific values which must be balanced. Ultimately, only the conscience of the individual scientist can decide.

A good example of this is Kuhn's account of the paradigm shift from eighteenth century Phlogiston theory to the modern paradigm in chemistry. When Phlogiston theory was abandoned in favour of the theory based on elements in the early nineteenth century, many older members of the chemistry profession resisted because they remained attracted by the wide degree of physical properties which Phlogiston theory was apparently able to explain. In contrast, the main attraction of the new theory

was that despite its limited explanatory scope, it was able to make predictions of mathematically precise proportions of ingredients in chemical reactions. Eventually, the attraction of the value of mathematical precision won the day, but it took more than another 150 years for developments in quantum mechanics to allow the new mathematical chemistry to explain the same breadth of physical features as the old Phlogiston theory.

The relevance of this analogy for the philosophy of economics is that the decision between different research programmes or paradigms in economics must also ultimately rest on values such as aesthetic appeal as well as sheer "number of facts explained" (which is of course a concept incapable of precise and rigorous operationalization anyway). Another factor which must be considered is the ethical dimension to the social sciences. Although there is an attractiveness to the view, espoused by both Kuhn and Popper, that there is nothing inherent in the social sciences which precludes them from having similar aims and status to those of the natural sciences, it should be admitted that if this view on the ultimately value-based nature of scientific discovery is correct, then the ethical image that the social sciences uphold for humanity must also be part of this value judgement. This links in with the indictment that interpretivist sociologists have made against economics that, to put it crudely, by developing its intellectual appeal, it helps to create a society of selfish egoists in its own image.

One does not need to accept the radical thesis that there exists no social reality independent of our theoretical constructs (whether through language, customs or social science) to accept that there is a great deal of validity in the idea that there is a feedback process between social reality and the concepts that social scientists use to explain and describe it. Economics has played an ethical role in promoting market societies, because of the view of many economists that individualistic societies produce greater economic efficiency and thus a better way of life than societies where people's economic behaviour is more closely controlled. From this perspective, the "core" assumptions of economics are not merely useful components of a predictive framework. They are, rather, central to the ethical vision of human nature, and its potential when permitted to develop freely, at the heart of economics. The view that economics should aim to be value free is thus self-defeating and self-deceiving. Pattanaik, for instance, has argued that value judgements have a necessary place in economic theory, and that, by elaborating the consequences of normative assumptions, progress can be made in normative theory in a rational manner, in a similar way to positive economics (Pattanaik, 1971).

Consider, for example, the assumption of rationality. In the normative sense that the preferences of the individual should be sovereign, this is not the kind of proposition that can be proved or disproved. The fact that people act as if they know what they want does not imply anything about the moral status of these "desires". The assumption of rational behaviour at the heart of many models in welfare economics is therefore much more than just a predictive modelling technique; it is the embodiment in economic theory of the moral value of a society based on respect for individual autonomy.

Despite its empirical usefulness, the assumption of selfish preferences does not, on the face of it, share the same positive ethical basis. This, arguably, provides a strong reason to bring the modelling of altruism into the heart of welfare economic theory. It also changes the emphasis of the level of analysis from that of empirical explanation of specific phenomena to that of assessing the economic efficiency, and therefore the social desirability, of different levels and forms of altruism in human societies. If we live in societies in which people exhibit partial altruism, or altruism in some contexts but not in others, then this is something that welfare economics must seek to explain, and not be content merely to assume. Indeed, it is probably the undesirability of making ad hoc assumptions about partial altruism that has so far led to the cleaner solution of simple self-interest remaining the main workhorse of abstract welfare economic analysis.

A model in welfare economics will require a number of properties if it is to satisfy the general prescription laid out above. One of the most important methodological questions which arises concerns the relationship between the social welfare function and the utility functions which the agents in the model seek to maximise. If the level of altruism is to be treated as an endogenous variable which can be altered (e.g. in different societies or via differential socialisation processes in the context of an existing society) then the possibility must be left open that each individual could be directly acting so as to maximise the social welfare function. Harsanyi has laid out the requirements for interpersonal comparability of utility if this benchmark case of perfect utilitarian, or "impartial", altruism is to be coherent (Harsanyi, 1986). (See section 1.5.) If we are to justify any kind of arrangement as being superior to this perfect utilitarian society, we need to introduce additional structure to the model (corresponding to Lakatos' peripheral assumptions). The opening quotation by Meade, for instance, suggests that there might be differences between the political and economic "marketplace" which justify a different level or kind of altruism as being socially optimal in different "scenarios" (Meade, 1973).

An interesting analogy to the role of the altruism assumption being suggested here is that of rational expectations in macroeconomics (Lucas, 1976). Just as there seems to be a kind of logical inconsistency between the assumption of rational agents with perfect understanding of an accurate model of the macro-economy and the use by these agents of adaptive expectations, there appears to be a parallel inconsistency between the assumption of moral human beings who design their society along the lines of utilitarianism but then act so as to selfishly maximise their own utility. Just as there are "hidden costs" to processing information which can explain why adaptive expectations are often a more plausible modelling technique in macroeconomics than rational expectations, there are "hidden costs" to individual altruism which may explain why it is socially optimal, despite the possibility of a society of utilitarian altruists within the structure of the model, for people to exhibit imperfections.

1.5 *Modelling Altruism*

Although altruism has become a somewhat peripheral issue in modern normative welfare economic theory, it has been an integral and well-recognised part of neo-classical economics for almost as long as the discipline has existed. Edgeworth and those who further extended his work quickly developed the implications of altruistic preferences for competitive equilibria in an Edgeworth box (Collard, 1978). The most important result from this body of work is what is known as the **non-twisting theorem**. This states that provided altruistic preferences are non-paternalistic, the contract curve will be the same as if individuals were fully selfish, except that it will be shrunken (the ends will be cut off) because extremely unequal distributions will be undesirable for both rich and poor. What is essentially required for this result to hold is that, in a precisely definable way, altruistic preferences should respect the relative valuations placed on different goods by the other individuals towards whom the altruism is directed. A formulation of social utility in terms of weighted sums of felicities will guarantee this. In the limit where individuals value each other as much as themselves, the contract curve shrinks to a single point. This is the bliss point for both individuals, since both individuals have the same utility function. This corresponds to the society of perfect utilitarian altruists introduced above.

The normative analysis of altruism raises some interesting additional issues. Suppose we have a society containing two individuals, each of whom cares about the other. As we have argued, it is necessary to distinguish between **felicity** (represented by $V_1(X_1)$ and $V_2(X_2)$), which is a measure of

the satisfaction that each individual gets from consuming goods, and **utility** (denoted by $U_1(X_1, X_2)$ and $U_2(X_1, X_2)$), which is a representation of the overall preferences of the individual as they determine his behaviour. Having made this distinction, we could take a number of directions in representing the two individuals' utility functions. If people care directly about each other's *utility*, then it is possible to get multiplier effects which can have perverse results. These will be discussed shortly. We could, on the other hand, have each individual's utility depend on a weighted sum of his felicity and the felicity of the other. Letting θ be the **coefficient of altruism**, and X_1 and X_2 be the consumption bundles of person 1 and person 2, this situation would be represented as:

$$U_1(X_1, X_2) = V_1(X_1) + \theta V_2(X_2)$$

$$U_2(X_1, X_2) = V_2(X_2) + \theta V_1(X_1)$$

The potential problem even with this formulation is that increasing the level of altruism automatically increases the utilities of both individuals, regardless of any effect upon their behaviour. The sequential punishment model presented in "The Socially Optimal Level of Altruism" (Chapter 2) nonetheless uses this formulation, because the felicities of the individuals in the model, rather than their utilities (which we shall call social utility, in order to be clear), form the basis of the normative assessment of the outcome of the model via a utilitarian social welfare function. This approach, however, does raise its own set of questions about whether it is legitimate to make a distinction between felicity and the "moral preferences" embodied in the social utility function. The justification is that we are seeking to assess moral preferences in terms of their contribution to economic efficiency. It could, however, be objected that this is an overly narrow view about the benefits of altruism, since compassion and empathy should be valued in and of themselves. The counter-argument, of course, is that compassion and empathy can lead to misguided actions, in which case it is hard to see why they should be automatically desirable.

Figure 1.2 shows a two person pure-exchange economy in an Edgeworth box where individuals are partially altruistic, with utility functions as specified below, and with $\theta_1 = \theta_2 = \frac{3}{4}$. Points c and d are the bliss points for the two individuals. Due to their partial altruism, and declining marginal felicity from goods X_1 and X_2, both would choose not to consume the entire endowment even if they could dictate the allocation (although, since they are only partially rather than fully altruistic, they would each like to take more than half).

$$V_1 = \sqrt{X_1} + \sqrt{Y_1}$$

$$V_2 = \sqrt{X_2} + \sqrt{Y_2}$$

$$U_1 = V_1 + \theta_1 V_2 = \sqrt{X_1} + \sqrt{Y_1} + \theta_1 \left(\sqrt{X_2} + \sqrt{Y_2}\right)$$

$$U_2 = V_2 + \theta_2 V_1 = \sqrt{X_2} + \sqrt{Y_2} + \theta_2 \left(\sqrt{X_1} + \sqrt{Y_1}\right)$$

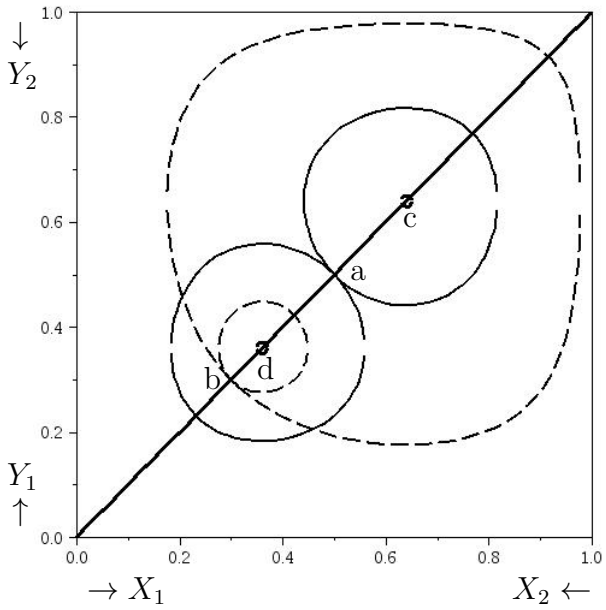

Figure 1.2: The shrunken contract curve

Point *a* shows a possible Pareto-efficient allocation, where there is a tangency between the two individuals' indifference curves. Point *b* also shows a tangency point, but it is *not* Pareto efficient, because moving closer to point *d* will make *both* players better off. The solid diagonal line represents all the possible points of tangency between the two individuals' indifference curves. However, the contract curve only consists of the part of this line lying between points *c* and *d*, because the two individuals would always bargain away from an allocation involving a level of inequality more extreme than *c* and *d*. Figure 1.3 shows the derivative of the individuals' marginal rates of substitution at each X_1 value along the tangency line. Only when these are both positive do we have Pareto-efficient allocations, since only then will the indifference curves "curve away" from each other in the standard textbook manner, where there is no altruism present.

This framework allows the introduction of one of the most commonly recognised problems with the translation from altruistic preferences to social outcomes. If the two individuals are indeed single

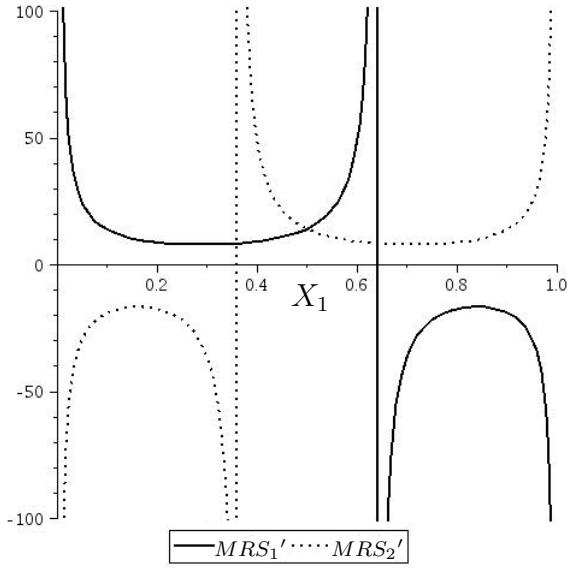

Figure 1.3: Derivative of the MRS of individuals 1 and 2

individuals, then it is reasonable to suppose that they would bargain to a point on the shrunken contract curve (i.e. to their unique shared bliss point if they are fully altruistic). However, if the two individuals in the model are taken to represent *groups* of individuals then there is a "public goods" problem in the sense that an individual in group A cannot redistribute to group B as a whole because a distribution of X to group B only increases the average income of group B by $\frac{X}{N_B}$ (where N_B is the number of members of group B). This implies that redistribution will not occur voluntarily. Hence a point like b can be a competitive general equilibrium, even though it is not Pareto-efficient.

A degree of state enforced wealth redistribution can, in theory, make both rich and poor better off in utility terms (essentially, greater equality is a state-provided public good). It should be noted, however, that this interpretation of group redistribution has already moved away from the assumption of perfect utilitarian altruism, since it has defined altruistic preferences as depending on the *average* utility of other individuals, whereas in a society of fully utilitarian altruists each individual cares for the marginal benefit of a gift to any individual *just as much* as that individual cares about himself. The public goods problem discussed above therefore does not arise in the case of full utilitarian altruism.

Although the alternative modelling approaches mentioned above have not been taken in "The Socially Optimal Level of Altruism" (Chapter 2), it is worth considering the consequences and issues that arise with them in order to achieve greater clarity. One possibility is to have each individual's utility be a weighted average of the felicities of all individuals (including himself). This approach

solves the problem which arose above of greater altruism automatically increasing the amount of utility in the economy. However, from the perspective of the sequential punishment model, it is an unnecessary complication. It would also be problematic because the model includes an infinite number of individuals. Although empathy with an infinite number of individuals in the sequential punishment model may seem unintuitive, it must be borne in mind that each individual need empathise with only one individual at any one time.

Another, more interesting, alternative approach is to have each individual's utility depend on his or her own felicity and the *utility* of the other individual. This could be represented as follows:

$$U_1(X_1, X_2) = (1-\theta)V_1(X_1) + \theta U_2(X_1, X_2)$$

$$U_2(X_1, X_2) = (1-\theta)V_2(X_2) + \theta U_1(X_1, X_2)$$

When this system of simultaneous equations is solved, we get:

$$U_1(X_1, X_2) = \frac{1-\theta}{1-\theta^2}\Big(V_1(X_1) + \theta V_2(X_2)\Big)$$

$$U_2(X_1, X_2) = \frac{1-\theta}{1-\theta^2}\Big(V_2(X_2) + \theta V_1(X_1)\Big)$$

Even though the utilities of both individuals have been normalised so that they are an average of their own felicity and the utility of the other individual, we still get a fairly complex multiplier term at the beginning of each individual's solved out utility function. This multiplier effect can produce interesting results in some models. Note however, that as long as certain regularity conditions are fulfilled (e.g. that $\theta \neq 1$) we can still express each individual's utility as being ultimately dependent only on the felicities of all individuals (and, of course, the coefficients in the model). This suggests that, if we want to abstract away from these multiplier effects caused by altruism, it would seem to be sensible simply to base individuals' social utilities directly upon felicities, as outlined above.

Bernheim and Stark have presented a model of marriage partnership which is a good example of the multiplier phenomenon in action (Bernheim & Stark, 1988). They assume that "females" are all equally altruistic, but that "males" vary in the degree of altruism towards their partner. Females choose male partners so as to maximise their own utility. It turns out that under certain conditions, females prefer a more selfish male partner because then there is less of a cost from the relatively low utility of the female entering into the utility of the male and then feeding back into the utility of the

female via the altruism of each partner for the other. Basically, women with difficulty getting utility from their own felicity (i.e. a low coefficient on their own felicity in their utility function) will prefer to be with a selfish male partner, who will not care that she is unhappy and whose happiness will perk her up at least a little bit. This provides a possible justification in economic theory for why "nice guys finish last" (although the coefficients in the model can sometimes lead the other way, with more altruistic males being preferred by females who are better able to get utility from their own felicity, due to the positive "spillover effect").

The definition of social utility functions in terms of weighted averages of felicities, and the assessment of outcomes using a social welfare function defined as the simple sum of felicities, requires ratio-scale interpersonal comparability of felicities (Roberts, 1980). As argued by Harsanyi (Harsanyi, 1986), only once we permit such interpersonal comparisons does it make sense to view the utilitarian social welfare function as requiring perfect, which he calls "impartial", altruism (all individuals weighted equally), and individuals' social utility functions as falling short of this by exhibiting imperfect altruism (lower weighting on some or all other individuals than upon onesself).

1.6 *Altruistic Punishment*

A major contemporary area of concern regarding the desirability of altruistic preferences is the role of punishment of defectors as a mechanism to achieve socially desirable outcomes. Empirical research by Fehr and his collaborators (Fehr & Gächter, 2000a) has shown that the ability of agents to punish, even at a cost to themselves, is of vital importance to prevent more selfishly inclined individuals from free riding. Fehr et al. have used small scale experimental prisoners' dilemma style games with real payoffs to demonstrate that people's anger against defectors leads them to be willing to altruistically punish, even when this comes at a cost to them so that it is individually irrational to do so.

The sequential punishment model presented in "The Socially Optimal Level of Altruism" (Chapter 2) examines the issue of punishment and the role it plays as an altruism amplification device from an abstract perspective. It is similar in nature to an infinitely repeated non-zero-sum stage game, with the added feature that the level of altruism is endogenous to the normative analysis of the model. The model demonstrates that there are offsetting benefits and costs to increased altruism. Although more altruistic individuals are less tempted to commit socially damaging acts, they are also less afraid of

being punished, because they care about their own utility less relative to that of others. Less altruistic individuals are also willing to punish the transgressions of others more frequently. It turns out that for a wide range of parameters, the level of altruism is irrelevant to the conditions required to achieve the socially optimal co-operative outcome. Under fairly general conditions, too high a level of altruism will even be counterproductive, and make a socially efficient outcome *harder* to achieve.

1.7 *The Evolution of Altruistic Preferences*

Treating the coefficient of altruism as endogenous to the optimization procedure encapsulated in the model obviously raises the question of how it is that a certain level or type of altruism comes about, and by what dynamic process this can be changed. This is not an orthodox area for economic modelling, which usually treats people's preferences as an a priori assumption rather than something which can be explained. There are, however, some precedents in the economics literature. There is also a vast literature in biology and cultural theory on the use of evolutionary theory to explain human altruistic behaviour. Usually, however, preferences are not explicitly modelled and instead highly simplified behavioural phenotypes are used.

It has been argued by Oded Stark that family demographics can be better explained through the passing of altruistic preferences from parents to children by example than by other forms of enlightened self-interest (Stark, 1995). One of the most interesting of the models discussed is one which seeks to explain why children's altruism towards their parents increases once the children have children of their own. Empirical evidence is cited to show that the presence of children increases the amount of altruistic behaviour by parents towards grandparents. Stark argues that this occurs so that parents can set the example to their children to look after them in a similar way when they become old. The microeconomic basis of such a framework is that there is a certain probability that children will simply copy the behaviour of their parents rather than do what is optimal for their self-interest.

The dynamic fragility of altruistic preferences has been reflected in models of the psychological processes which lead to norms of co-operative altruistic behaviour and the potential for government intervention to negatively impact upon them, resulting in possible welfare loss (Hollander, 1990). Work has also been conducted on models of mutually reinforcing altruism among individuals in situations of strategic complementarity such as the workplace (Rotemberg, 1994). Individuals whose preferences

become genuinely more altruistic may do better if this leads others to be more altruistic towards them. Other approaches to the dynamic survival of altruistic behaviour have emphasised the connection between altruism and "docility" due to limited human cognitive abilities (Simon, 1993). The idea here is that because individuals cannot distinguish between cultural norms which improve their own fitness and those which require altruistic behaviour, altruism can "piggy-back" on the back of other norms provided that the overall impact is to improve individual fitness.

1.8 *The Multilevel Selection Paradigm*

The sociobiological approach would perceive differing levels of altruism as the result of differing levels of genetic relatedness. However, as we have seen, this does not seem to be adequate in explaining the role of cultural factors in determining the altruistic interactions between genetically unrelated individuals which is the vital glue that holds together a complex society, regardless of the degree to which market forces are relied upon to achieve economic coordination. The functionalist approach in sociology, most famously associated with the work of Robert Merton, would seek to explain the presence of a particular level and pattern of altruism in terms of its functionality for the overall social system (Merton, 1968). This approach, however, can be dangerous if it fails to use a modelling approach based upon methodological individualism and if it neglects to provide a causal explanation to underpin the functionalist one.

More recently, the social sciences have seen sophisticated and innovative attempts to use Darwinian ideas to integrate the functionalist and individualist approaches to the study of human society (Sober & Wilson, 1999). It has been argued that evolutionary progress has resulted in the functional integration of smaller units into larger ones via the process of "group selection". For example, competing strands of DNA "work together" within cells, as do cells within the body. There seems to be no good reason to assume that functional integration on the societal level should be impossible, and thus individual human animals can be seen as working together cooperatively within society in an analogous way.

The multilevel selection approach, however, demands that the explanation for the emergence of group-functional behaviour also takes into account the potential for conflict between the component members of groups. For example, sometimes "selfish genes" will take actions that lower the fitness of the other genes in order to benefit themselves. However, the structure of a DNA strand makes it

fairly difficult for individual genes to do this. Dawkins uses the analogy that they are like "rowers in a boat" who are unable to escape their dependence upon each other and so are forced to work together in order to maximise their own survival chances (Dawkins, 1976). The situation is compatible with an individualist explanation[3] of the functional integration of individual animal bodies because genes are better off, overall, as part of DNA strands protected within cells and bodies, than as separate molecules in the outside world. Groups of genes whose "boat" does not tie them together sufficiently firmly will not succeed in replicating themselves.

The analogy between genes and DNA strands works also between individual bodies and social structures; social structures, like the organization of genes into DNA strands and cells, help individuals to overcome their individual selfishness and work together to improve their overall fitness. In this sense, human society as a totality acts as an altruism amplification device by providing a structure in which the temptation towards socially harmful acts of individual selfishness is reduced. The existing types of human social organization must be explained partly by the process of competition between groups leading to the expansion of more efficient forms of organization and partly by the need for any social system to contain destructive competition between members, which may require that certain features be present, the punishment of wrongdoers being a particularly important example.

The "invisible hand" of the market, as rigorously encapsulated in the First and Second Theorems of Welfare Economics (Arrow & Debreu, 1954), can also be understood within this context as showing how, provided a complete set of markets is created with well-defined property rights, entirely selfishly motivated economic activity *within* this system leads to a mutually beneficial outcome in which all individuals are made better off than their original endowment. This suggests that social control of all economic activity[4] is not necessary to achieve efficient functional integration at the level of human society. On the other hand, the temptation of individuals to break the law for their own gain will never be fully removed due to the evolutionary pressure against functional integration at the individual level, and so the idealised assumptions of the First and Second Theorems are unlikely to be fully achieved.

Economic theory has other important roles to play in this Darwinian framework because it provides a set of tools to assess the efficiency of particular institutional arrangements through use of the "ideal benchmark" of a social welfare function and paradigms for explaining theoretically the success or

[3]Individualist in the sense of starting from the selfish gene as the fundamental explanatory unit.

[4]This could be achieved through direct state control or by, for example, a very strict moral code.

failure of decentralised social systems to achieve efficient outcomes. Normally, the preferences of individuals are assumed to be fully selfish and other features of the model such as the presence of asymmetric information or externalities are used to explain the differences in the performance of market coordination of economic activities in various situations. No reason, however, seems to preclude treatment of the level of altruism present in individual preferences as an alterable variable that may have an impact upon social efficiency. The sequential punishment model presented in "The Socially Optimal Level of Altruism" (Chapter 2) takes this approach.

1.9 *The Limits to Altruism*

There are three logically distinct facets to the concept of limits to altruism. Firstly, there are limits to how much there is in society. We shall call these "positive limits". We shall explore these in sections 1.10 to 1.16. Secondly, there are limits to how socially beneficial altruism can be. As laid on in the road-map in section 1.2, there is a key distinction to be made here between particular kinds of altruistic action, and altruistic preferences themselves. We will focus on these "normative limits" in section 1.17. Thirdly, there is the question of the broader consequences to the presence of these limits, particularly for normative policy issues. We explore these in section 1.18

1.10 *"Stylised Facts" About Altruistic Behaviour*

The idea that human beings are motivated partly by altruistic and partly by selfish objectives, and that individuals differ in the degree to which they exhibit these types of preference, is fairly intuitive and non-controversial. Explaining why this pattern should emerge is more difficult. Thus far, we have considered the overall role that economic theory can play in helping to explain the altruistic behaviour we observe. We will now consider evidence from empirical economics for the "stylised facts" about the imperfections of real world altruism that the models presented in sections 1.17 and 1.19 are needed to help explain. These can be summarised as:

(1). Altruism is often present as a motivating factor in the economic realm.

(2). Altruism is frequently subject to imperfections.

(3). These imperfections differ between individuals.

(4). Malevolence is frequently evident.

Our aim in the following sections will be to substantiate the above statements. Before considering the empirical evidence in detail, it is helpful to quickly summarise the areas of empirical economics where evidence of altruism can be found. Firstly, there are results of experiments where individuals play games with small payoffs (usually, but not always, in cash form), some of which have already been mentioned. Secondly, there is the evidence of individuals making charitable donations and contributing voluntarily to the provision of public goods. Closely connected, there is thirdly the issue of parental bequests to their offspring. A fourth body of evidence comes from responses to surveys about willingness to pay for certain public goods, such as environmental protection. There is, finally, evidence of limits to altruism from issues of international political economy.

1.11 *Experimental Economics*

One of the most obvious areas of empirical research in economics that provides evidence of the failure of the assumption of rational self-interested behaviour involves the use of experimental games with small (usually monetary) payoffs. The classes of game most commonly played are the ultimatum game, the public goods game, the prisoners' dilemma, the centipede game and the bargaining game. In all these cases, it is well established that the observed behaviour conclusively violates the predictions of the standard theoretical framework based on rational self-interest. However, since self-interest must always be combined with other assumptions, such as full knowledge of common rationality, it is always open (but with varying degrees of plausibility) to consider these peripheral assumptions as having been falsified, rather than self-interest itself.

1.11.1 The ultimatum game

The "ultimatum game" is played between two individuals. The first individual proposes the division of £1 between the two individuals and the second individual can either accept the offer or refuse, in which case both get a payoff of 0. If both individuals are entirely rational and selfish, and there is full common knowledge of rationality, non-co-operative game theory would predict that the first individual will offer the smallest amount they can that is higher than 0 (i.e. 1p) and that the second individual will accept. However, when the game is played in experimental situations with real people, the predicted outcome occurs extremely rarely, and there is significant variation between cultures

regarding the amount that the first individual offers to the second.[5] It is clear, therefore, that in most cases individuals are constrained from fully selfish behaviour by moral norms. Not only are people in the position of individual 1 observed to offer more than the minimum to the second person, if the offer is not generous enough, people in the position of individual 2 are observed to refuse the offer, even though this is economically irrational for them.[6]

The research into the ultimatum game and some variants has revealed a number of other important phenomena. Camerer and Thaler also cite experiments comparing the ultimatum game to the dictator game, where the second individual is simply forced to accept the allocation proposed by the first individual (Camerer & Thaler, 1995). In the dictator game, offers are usually lower than in the ultimatum game, but not as low as predicted by pure self-interest. This shows that the offers in the ultimatum game are being partly driven by a desire of the proposer to be fair and partly by fear that the second individual will refuse. Other experimental results referenced in the same paper show that changes in the framing of the question (which should be irrelevant to the strategies played if individuals are self-interested), such as whether the proposer is chosen randomly by a lottery or arbitrarily by the researcher, and whether or not the income is described to have been "earned" by the proposer, can make a significant difference to the altruism exhibited by the proposer towards the second individual.[7]

Andreoni and his collaborators (Andreoni et al., 2003) have extended the ultimatum game to convexify the strategy space of the second individual by allowing them to continuously shrink the "pie" after the allocation is chosen by the proposer (the standard ultimatum game allows a discontinuous choice only between 100% and 0% in terms of the possible moves by the second individual in this convexified ultimatum game). This has the advantage that the experimental data becomes sufficiently rich to allow the axioms of revealed preference to be tested. They find that the behaviour of experimental subjects is compatible with the hypothesis of rational altruism.

Individuals have preferences not just over their own money payoff, but over the money payoff of the other individual. These preferences can be both "benevolent" and "malevolent" in the sense that if an individual has a high monetary payoff relative to others, then they gain utility when others get a greater monetary payoff, but if the other individual is already better off in monetary terms then

[5] The empirical evidence has been summarised (Camerer & Thaler, 1995) as showing that offers are usually between 30% and 40%, with the mode often being 50%.

[6] Very few offers are below 20%, and those which are this low are often rejected (Camerer & Thaler, 1995).

[7] Clark finds similar evidence for framing effects in a game where individuals can make costly votes to reduce outcome inequality (Clark, 1998).

the individual's utility is *decreasing* in their opponent's monetary payoff. These types of rational inequality-aversion explain why it is rational for individuals to make more equal offers at their own material expense and to "refuse" (in the sense of partially or fully shrinking the pie) unequal offers, again at their own material expense. Andreoni finds a very similar result with experiments using the dictator game (Andreoni & Miller, 2002).

1.11.2 The public goods game

The public goods game involves a situation where a number of individuals must choose whether or not to contribute to a public good. Each unit of contribution produces more than 1 unit of the public good. This is split over the N individuals, however, and so it does not pay each individual to contribute if they are self-interested, because they are still able to free ride on the contributions of other individuals. The evidence has been summarised (Dawes & Thaler, 1988) as showing that contributions are usually in the region of 40%-60%.[8] When the game is repeated with the same individuals playing, the average level of contributions tends to drop over time. Fehr and Fischbacher argue that this can be explained by a model in which the desired level of contributions depends upon the amount others contribute[9], but that desired contributions climb at less than a slope of one, either because there are selfish individuals present or because altruistic individuals are not completely willing to match the contributions of others (Fehr & Fischbacher, 2003). This causes the levels of contribution to the public good to "unravel" over time.

Framing effects have also been found to be significant in public goods games. It has been found (Andreoni, 1995b) that the level of contributions is significantly affected by whether the public good is seen as the provision of a positive benefit or the avoidance of a bad externality, even though the payoff structure is identical in both cases. Such results could only be explained if individuals have utility functions which value the size and structure of changes to outcomes, as well as final outcomes. Other research which shows that allowing group discussion before contributions are made greatly increases the contribution level (Dawes & Thaler, 1988). This may be because people are induced to make promises to contribute, which they then feel are binding, or because they get to know the members of their group, and thus feel more altruistic towards them. Interestingly, when individuals are split into

[8] Andreoni considers whether this is due to imperfect rationality or altruism, and concludes that both play a part (Andreoni, 1995a).

[9] This would require some kind of altruistic preference for reciprocity ("if others contribute then so should I").

groups but the beneficial externality goes to another group, the result is reversed and free riding is reinforced because the group agree together to act in their group self interest and refuse to contribute to the intergroup public good. This is strong evidence for the narrow-group-based nature of human co-operation which, as we shall see later, relates closely to explanations of the evolution of altruism based on group selection.

1.11.3 The prisoners' dilemma

The main empirical anomaly for the self-interest assumption that concerns the prisoners' dilemma is the observation that co-operation occurs in a finitely-repeated prisoners' dilemma game when the well-known backwards-induction argument should lead co-operation to unravel and defection to occur from the first period onwards. Andreoni and Miller conclude on the basis of experimental evidence that rational reputation-building on the part of most agents plus true altruistic behaviour on the part of a minority offers the best explanation for this phenomenon (Andreoni & Miller, 1993).

1.11.4 The centipede game

The centipede game is a sequential-move game with finite length in which two players get chances in sequence either to take the larger of two payoffs on the table, or to pass, in which case both payoffs are multiplied by a factor. After N alternating moves ($\frac{N}{2}$ for each player), the game ends. By backwards induction, the last player to move will take the larger pile, since otherwise they get nothing. Since the payoffs are set up so that the larger pile at the $N-1^{\text{th}}$ move is bigger than the smaller pile at the N^{th} move, the penultimate player to move will take the larger pile. Therefore the player at the $N-2^{\text{th}}$ move will also take the larger pile. By backwards induction, this reasoning can continue to be applied to show that the first player to move will take the larger pile, even though if the two players could co-operate they could end up with massively higher payoffs.

McKelvey and Palfrey conduct experimental centipede games and find that typically players pass for a number of periods before somebody takes the larger pile (McKelvey & Palfrey, 1992). They explain this using the idea that a proportion of the population is altruistic, and that selfish individuals can pretend to be altruistic in order to get their opponent to co-operate. By calibrating the model to their data, they estimate that 5% of the population is believed to be altruistic.

1.11.5 Bargaining games

Unlike the other games, bargaining games do not have sufficient structure to make unambiguous equilibrium predictions. Assuming individuals are selfish, however, co-operative game theory does unambiguously predict that no player will bargain to an outcome that is worse than their status quo payoff. Hoffman and Spitzer have used experimental bargaining games to show that this prediction is violated (Hoffman & Spitzer, 1985). Given a chance to work together in a co-operative endeavour, individuals seem willing to share payoffs equally even when their outside options are unequal.

1.11.6 The nature of human altruism

Ernst Fehr and a number of collaborators have recently worked on a synthesis of the results of experimental economics with evolutionary theory (Fehr & Fischbacher, 2003) (Fehr & Gächter, 2002a). They argue that empirical results such as the ones outlined above can only be explained by the presence of what they call strong reciprocity in human motivation (Fehr & Gächter, 2000b). Weak reciprocity is the form of altruism that can be seen as enlightened self-interest[10] – individuals do each other a good turn because they rationally expect to be "paid back". However, this limited form of altruism is not sufficient to explain the complex functional integration of human societies consisting of insufficiently genetically related members, and is also insufficient to explain the observed experimental regularities.

Strongly reciprocal altruism can take a positive and negative form, where individuals either help or harm others at true material cost to themselves. This acts as the "glue" that holds social institutions together, because the willingness on strong reciprocators to punish cheats even at cost to themselves forces selfish individuals to also behave themselves. Fehr and Gachter find that the ability of players to make costly punishments significantly increases the level of co-operation in the public goods game (Fehr & Gächter, 2000a).

The importance of the willingness of individuals to engage in altruistic punishment has also been reflected in recent work on cultural selection theory. Altruistic punishment is the main mechanism by which social organization can act as an altruism amplification device, because it is usually less costly to punish another individual (e.g. by ostracising them) than it is to make an altruistic sacrifice for their benefit (Sober & Wilson, 1999).

[10]Category (2) in the road map.

1.12 *Voluntary Giving*

The empirical result that the rational self-interest assumption does not correctly predict subjects' behaviour in experimental public goods games, along with the casual observation that people do in fact contribute en masse to public goods using both their time and their money, has led to attempts to develop microeconomic models of charitable giving that fit real human behaviour.

Andreoni constructs a model of contributions to a public good, where individuals must choose how much of their wealth to donate to the public good, with the rest being privately consumed (Andreoni, 1990). He shows that a pure altruism model, where individuals care only about their consumption of the private good and the overall amount of the public good provided, fails to explain important empirical phenomena. A pure altruism model predicts that forced donation to the public good (e.g. through taxation) should be completely crowded out by reduced contributions, because the individual's budget constraint, given expected contributions by everyone else, will remain unchanged, and so the optimally chosen combination of private and public goods consumption will also stay the same. The evidence, however, shows that crowding out is much smaller than this.

A pure altruism model also predicts that, assuming everyone has identical wealth and preferences, subsidising voluntary donations by a certain amount will be identical to taxation because the subsidy per unit of donation will be removed from the representative individual's wealth level, leaving the opportunity cost of donating unchanged. This renders the policy of subsidising donations unjustified.

By replacing the pure altruism model with a model of impure altruism where individuals also care about their specific donation, the empirical regularities can be more adequately explained, and the subsidization of private giving can be shown to result in a greater increase in contributions than the equivalent direct expenditure on the public good from taxation. Sugden similarly argues that pure altruism cannot explain the observed behaviour of contribution to public goods (Sugden, 1982).

Frank et al. find evidence for the theory of warm glow as opposed to pure altruism in data on charity care by private nonprofit hospitals in the US (Frank et al., 1996). McGranahan finds evidence from 17th century wills which supports the existence of a combination of altruism and warm glow motives for charitable bequests (McGranahan, 2000). Freeman studies volunteer labour supply and finds results compatible with the warm glow hypothesis; individuals are discerning about the causes they support and make rational decisions whether to donate time or money (Freeman, 1997).

Sugden explicitly models reciprocal preferences by considering individuals who follow the moral norm that "if everyone else provides at least their fair share in my eyes then so should I" (Sugden, 1984). This turns the public goods provision game from a prisoners' dilemma into an assurance game, where there will be multiple equilibria but only one of which will be Pareto efficient. This is an impure form of altruism because people's perception of a fair contribution depends upon their valuation of the public good, whereas a pure altruist would value it at the sum of the valuations of all individuals.

1.13 *The Economics of the Family*

1.13.1 The "rotten kid" theorem

The application of economic methodology to the study of the family was pioneered by Gary Becker. The most famous result to emerge from his work is the "Rotten Kid Theorem" (Becker, 1974) (Becker, 1981). This could be thought of as a theory of how the family can act as an altruism amplification device. The theorem states that all members of a family will behave efficiently, even if they are completely selfish (or imperfectly altruistic) provided that the head of the family is sufficiently altruistic to make an operative transfer. (The head of the family is defined as the member who has sufficient private income to make positive transfers to all other family members due to altruism.) The head of the family thus provides an "altruistic linkage" between all other members.

Suppose we examine the decision of a "rotten kid" (selfish family member) over whether to take an action that increases or decreases their pre-transfer income at the expense or gain of another family member. The head of household will take into account the decision made by the rotten kid when deciding how big a transfer of resources to give him. In other words, the head of the family ensures, through the alteration of the size of the transfer, that the rotten kid gets a share of the total family wealth as determined by their need relative to other family members, as perceived by the family head. Any action which increases the overall collective family wealth therefore makes the rotten kid better off. The rotten kid thus behaves fully efficiently; the operative altruistic transfer from the family head ensures that they fully internalise the costs and benefits their actions cause to other family members.

The technical conditions required for the Rotten Kid Theorem to hold were clarified by Bergstrom, who concluded that, although not universally applicable, it is still valid in many situations (Bergstrom, 1989). The requirement is that the actions taken by the rotten kids must be of such a nature that they

cause a shift rather than a change in the slope of the family's utility possibilities frontier. This ensures that the family head, provided she is non-paternalistic and benevolent towards all family members, will definitely reallocate income so that the rotten kid is better off individually when he takes an action which expands the family utility possibilities frontier.

When rotten kids can take actions that distort the slope of the utility possibilities frontier, the rotten kid theorem no longer applies and they may take inefficient actions from the perspective of the family. It has been pointed out (Bruce & Waldman, 1990) that one of the main situations in which this condition does not apply is when there is a moral hazard problem between the head of the family and the rotten kid due to certain goods in the rotten kid's utility function not being under the direct control of the head of household via the transfer process. This is also known as the "Samaritan's Dilemma". Andreoni has described the problem as that when parents can be manipulated by children into "spoiling" them then they do become "spoilt rotten" (Andreoni, 1989).

1.13.2 Evidence on transfers and "altruistic linkage"

The Rotten Kid Theorem emphasises the importance of transfers between family members. One of its main empirical predictions is that redistribution of income between family members should not affect the distribution of consumption. This prediction has been tested on data on the US extended family (Altonji et al., 1992) and the Japanese extended family (Hayashi, 1995). Both studies found conclusively against the hypothesis that extended families are "altruistically linked" by a single head. Although this does not necessarily show imperfect altruism on the part of individuals within an extended family, since there are other factors such as asymmetric information and non-operative transfers in play, it is at least suggestive.

Empirical evidence (Altonji et al., 1997) also shows that the theory of pure altruistic linkage cannot explain other phenomena such as the responsiveness of inter-vivos transfers to changes in parental and child income. Altonji et al. argue that introduction of a "warm glow" factor into the model could allow this to be better explained. Bernheim and Bagwell have argued that the central flaw with the traditional model of the altruistically-linked dynastic family is the single-linkage assumption, because, once multiple linkages are recognized as part of "dynastic networks", such a framework has absurd implications which are clearly empirically false (Bernheim & Bagwell, 1988).

The altruism-based model of transfers within the family has also been challenged by the "exchange" approach, which sees the family as an institution enabling insurance contracts to be made between self-interested members. It is difficult to settle the issue empirically due to the similarities in the predictions of both frameworks (Kotlikoff & Spivak, 1981). Some studies have found that exchange is a more important explanatory factor than altruism (Cox, 1987) (Cox & Rank, 1992). Altruism has still often been found to improve the explanatory power of models which explicitly take into account such possibilities, however (Tcha, 1996) (Sloan et al., 2002). It has been argued that the altruism, exchange and warm glow models should not be seen as mutually exclusive and should be integrated together in a pragmatic manner (Stark & Falk, 1998).

1.13.3 Bargaining models of the family

The empirical failure of models based upon the notion that decisions within the family can essentially be modelled as being made by a single head has led to a body of work applying bargaining theory to decisions within the family. The most commonly used framework is that in which a husband and wife with differing interests bargain over the allocation of resources within the family. It has been found that although the empirical evidence rejects the hypothesis of decisions being made in accordance with a single household utility function, it does fit the hypothesis of efficient bargaining between members with differing, even if partially altruistic, utility functions (Chiappori, 1992) (Browning & Chiappori, 1998). This therefore constitutes strong evidence for imperfections to altruism even within the nuclear family. Research on divorce has found evidence that such efficient bargaining breaks down, resulting in a reduction in child welfare (Weiss & Willis, 1993) and the idea of bargaining between men and women has also been used to explain out-of-wedlock childbearing (Willis, 1999).

1.13.4 Biased altruism

Some of the most interesting evidence on the imperfections to altruistic motivation within the family comes from the study of unequal treatment of children on the basis of gender, birth order and biological relatedness. Researchers into differentials in wages and human capital investment between males and females in the US (Behrman et al., 1986) have attempted to determine whether or not this is driven by greater weight upon the success of male children in parents' altruistic utility functions. The conclusion of this study was that existing wage differentials reinforce inequalities in human capital

investment, but that parents do not, at root favour boys (in fact, if anything, the raw weighting on the welfare of girls is slightly higher). A similar study of the Phillipines (Davies & Zhang, 1995), however, found evidence for pure gender bias in parental preferences, underlining the fact that altruistic imperfections are culturally relative.

Research has also been conducted into why first-born children tend to do better in terms of measures of life success than children born subsequently to the same parents (Behrman & Taubman, 1986). The conclusion is that this is due to natural endowment effects rather than parental favouritism (as with gender bias, it may again be the case that parents compensate by weighting later-born children slightly *more*). Evidence on the treatment of step children in the US (Case et al., 1999), on the other hand, suggests that they do receive a smaller proportion of family income on food if they live with a stepmother, after controlling for income.

1.13.5 Discount rates

Research into the discount rates parents are revealed to apply to their children's welfare in making costly decisions related to children's health, as measured by lead contamination, has shown them to be similar to market interest rates for wealthier parents and higher for poorer parents, but nowhere near as high as the discount rates of 20% - 50% found to be applied to consumer durables (Agee & Crocker, 1996). This suggests a strong degree of parental altruism towards children.

1.14 *Bequests*

One of the most important topics concerning altruism in economic theory is that of bequests, most often from parents to children. There are two dimensions to this issue. The first is the microeconomic question of what motivates people to leave bequests. The second is the macroeconomic question of the implications of bequests for the role of government debt, income inequality, savings and investment and economic growth. Evidence in both these areas is potentially relevant to the existence of altruism, and its imperfections, because the answers to questions in the dimension of motivation have macroeconomic implications, and vice versa. There is a massive body of empirical research in this area, so we will only be able to take a cursory look at the main issues and findings.

Bequests have been estimated to account for about 80% of net private wealth in the US (Kotlikoff, 1988). A number of potential explanations have been offered for why parents leave bequests.

It may be due to a genuine altruistic concern for children's well-being. Such altruism may be of the "pure" or "warm glow" kind, as we will discuss in further detail below. Bequests may, on the other hand, be accidental, as a result of the parents dying earlier than they expected, and thus being unable to consume their entire wealth. A more sophisticated variation on this theme, which can potentially explain a greater role for bequests, is the theory that children and parents enter into a contract that parents will leave positive net wealth in return for children supporting them if they live longer than expected.

The microeconomic evidence for the importance of altruism in determining bequests is mixed. However, a reasonable conclusion would seem to be that there is genuine but imperfect and heterogeneous parental altruism towards children. Studies have found that inter-generational altruism is a plausible explanation for observed bequest behaviour (Adams, 1980), that individuals value bequests as highly as their own consumption[11] (Kuehlwein, 1993), that bequest behaviour reflects a high degree of heterogeneity, arguably driven by differences in preferences, even among a high income US sample (Laitner & Juster, 1996), and that there are statistically significant differences in bequest behaviour due to differences in income level, race and head of household gender (Kurz, 1984).

A pure altruism model of parental bequests is problematic. If parents are directly concerned with child utility, they should give different bequests to different children based on the children's income. Empirical evidence, however, does not support this prediction (Wilhelm, 1996). Parents tend to give equal bequests regardless of their children's differing situations. This suggests that some kind of impure "joy-of-giving" model where it is the size of bequest that enters directly into the parental utility function would be more realistic. It has been shown (Abel & Warshawsky, 1988), however, that in terms of the overall macroeconomic implications, the two theories are interchangeable in that empirically observed joy-of-giving parameters imply imputed coefficients of altruistic preference for children's welfare that are not implausibly high.

The resistance to accepting the primacy of parental altruism as the cause of bequests comes primarily from its macroeconomic implications rather than from the microeconomic sphere. In a seminal paper (Barro, 1974), Robert Barro showed that if parent and children are altruistically linked into a "dynasty" by a planned positive transfer of wealth (an operative transfer) either from parent

[11] Interestingly, this was even found to apply to individuals without direct heirs, providing evidence for more generalised altruism also.

to child or from child to parent, then cutting taxes and issuing government debt will have no effect on consumption expenditure because it does not alter the dynastic budget constraint out of which the altruistic parent (or child in the less frequent case of a child making the operative transfer to a parent) distributes income. This is the famous Ricardian Equivalence Theorem, which implies that government debt does not either crowd out private investment or cause a trade deficit. The traditional Keynesian view, however, is that government debt does crowd out investment and net exports. Although more recent attempts to weight the evidence have questioned the empirical rejection of Ricardian equivalence (Seater, 1993), other strong neutrality results not borne out by empirical evidence can also be shown to hold if the altruistically linked dynastic model is true, such as the invariance of inter-temporal consumption to inter-generational redistributions of income. (Kotlikoff, 1988).

It is thus probable that not all families are dynastically linked, because they do not have sufficiently strong altruism towards their children to make the transfer operative. (Note that, as argued by Bernheim and Bagwell, the observation that some people do not have children is not a good basis for an argument against Ricardian Equivalence because, going back far enough, all human individuals are related (Bernheim & Bagwell, 1988).) These are, of course, only a few among many reasons why these neutrality results are unlikely to hold fully in the real world. The use of the altruistic bequest motive in overlapping generations policy simulation models has now become commonplace (see, for example, (Altig, David et al., 2001)) and the consensus view would seem to be that it provides the best fit for building realistic models (Rangazas, 1996) compared to alternative theories.

1.15 *Environmental Goods*

Environmental goods are one of the most commonly examined types of public good for which there is evidence that individual valuations depend to some degree upon altruistic preferences. Much of this evidence comes from the use of the contingent valuation methodology, which asks survey respondents to give a valuation of a particular environmental good in a particular context. This method can be justified (Hanemann, 1994) on the grounds that its results have been found to match up reasonably well in specific cases with the findings of the traditional revealed preference approach. Hanemann concludes that although contingent valuation is subject to potential biases and inaccuracies, so also is revealed preference.

A study of voting behaviour in a Californian referendum on tightening water quality regulation (Holmes, 1990) found evidence for altruistic as well as selfish motivation for voting behaviour. A proxy was constructed for an environmental altruism variable using the residuals from regressing the results of an earlier Senate race, fought between a Democrat with a strong environmental voting record and a business Republican, on variables to proxy partisanship and a vector of self-interest variables. It was found that this proxy variable had a statistically significant impact upon voting in the referendum. This suggests an independent role for altruistic preference in contingent valuation as evidenced in voting behaviour, although the effect would be reduced and possibly eliminated if some variables were missing from the vector of self-interest variables.

Research on the relationship between environmental valuations reported in survey data and life expectancy (Popp, 2001) has also found evidence of a role for partial altruism. If people are fully altruistic, their life expectancy should not affect their contingent valuation of environmental goods. On the other hand, if they are fully selfish, the valuation should be, on average, zero as life expectancy goes to zero. The evidence, however, rejects both these hypotheses, suggesting the presence of partial altruism. The central estimate is of an equal weighting between individual welfare and the average welfare of future generations, but this estimate is very sensitive to the assumed discount rate.

Weak evidence has also been found (Weaver, 1996) of a role for altruistic valuation of environmental goods in *production* as well as consumption decisions. The decisions of Pennsylvania farmers to invest in particular technologies were found to depend upon the valuation of environmental goods as well as narrow profit motives.

The precise manner in which additional factors over and above self-interest drive the contingent valuation of environmental goods remains unclear. Some evidence indicates that the altruism here is connected with the desire of individuals to make a "fair" contribution rather than being driven by pure altruism (Stevens et al., 1994).

1.16 *International Comparisons*

Some of the most striking evidence for imperfections to altruism comes from the realm of international political economy. A 1998 study which sought to estimate the marginal cost of additional life expectancy in different countries found that the implicit valuation of a life year in the richest

countries was 300 times that in the poorest countries. Once difference in average life expectancy are taken into account, the cost of saving an entire lifespan in the richest countries came to 1000 times that of the poorest countries (Dowrick et al., 1998).

One of the most documented phenomena in comparative political sociology is the difference in the extent of the welfare state between the US and the EU countries.[12] It has been argued (Alesina et al., 2001) that the most convincing explanation of these differences lies in differing attitudes to the poor, with Americans being more likely to view the less well off as lazy and responsible for their own situation.[13] This was demonstrated by regressing responses to an attitude survey on various possible causal variables. It was found that the dummy variable on the US remained statistically significant, demonstrating "American exceptionalism". The authors of this study tentatively suggested that racial fragmentation may be the causal factor, since the inclusion of a racial fragmentation index resulted in the dummy variable on the US becoming statistically insignificant.

1.17 *Normative Limits*

The volume of mainstream economics literature provides a number of powerfully suggestive reasons why altruism (of the fully utilitarian rather than "average utility" form) will have costs as well as benefits to society. These come about through the abandonment of the stringent assumptions of the non-twisting theorem (essentially also those of the perfect competition general equilibrium model).

One area that has been given extensive consideration is the compatibility of profit-making behaviour with altruistic motivation. Is an altruistic society necessarily a non-market one, and is it necessary for individuals to be selfish for a market society to operate efficiently? Becker has made the point that if owners of firms are altruistic towards other individuals, it is more efficient to give money away to the poorest than to subsidise the price of goods below marginal cost (Becker, 1981). This of course assumes that the market in which the altruistic firm operates does not in the first place suffer from market failures such as externalities, which would require self-imposed "subsidies" or "taxes" of a fully altruistic firm in order to internalise the externality.

[12] General government spending makes up 36% of GDP in the US, but an average of 48% in the European Union. Subsidies and transfers make up 11% of GDP in the US and 20% in the EU, thus accounting for three-quarters of this difference. The primary difference would seem, therefore, to be in the redistributive role of the state.

[13] A similar study (Freeman, 1984) found that differences in "personality type" in different countries helped to explain the size of the welfare state.

Even in the presence of market failure, the desirability of acts of "corporate social responsibility" also depends heavily upon informational considerations. Altruistically motivated behaviour requires that agents know the preferences of others, so that they are able to act so at to make them happier. A number of economists (Baumol, 1975) (McKean, 1975) have argued that whilst moral systems based on altruistic motivations work well in the case of maxims like "Do not be rude to people", they are unlikely to work in the case of rules such as "Do not produce too much pollution" because the issue of how much pollution is socially optimal depends on information which simply is not available in such a form as to elicit a social consensus, even among a society of perfect altruists. Thus some issues should be left to regulation through the democratic system rather than unilateral attempts to act in an "ethical" manner by firms. The democratic system will be better able to predict, and thus regulate, firm behaviour if it can expect that they will act in a clear profit-maximising manner.

Limitations to the potential for altruistic socially responsible behaviour also apply to individual consumers as well as firms. For example, take the issue of the choice to buy organic foods. Assume firstly that these are more expensive than the non-organic versions and secondly that the felicity from consuming the two alternative products is the same. Assume also that both markets are competitive and operate in the same regulatory system so that there are no issues of monopoly or exploitation of labour. Compare two individual policy options. (1) Buy the more expensive organic product. (2) Buy the cheaper normal version and donate the relative saving to the poor. If the organic product is more expensive because it reduces a harmful externality with value in excess of the price difference, then there is a trade-off between the more efficient option (1) and the more equitable option (2). If, on the other hand, the organic option simply uses more resources (e.g. land, labour time) then option (2) is superior on both efficiency and equity grounds. It may be very difficult for a consumer to have the necessary information to make the optimal decision.

There are normative limitations not just to particular manifestations of altruistic behaviour, but to altruistic preferences themselves. A major possible drawback from altruistic motivation is that it may serve an antisocial purpose by helping specific groups to achieve rents via collective action (e.g. business lobbies, labour unions, cartels). As the application of the economic theory of public goods to political collective action shows (Olson, 1965), groups whose members feel more altruistic towards other members of the same group will be more successful at overcoming the free rider problem in their

political collective action, and will be able to extract rents from the rest of the population in a socially inefficient manner. If it is very difficult to achieve generalised other-regarding preferences (as seems likely) then it may be better for people to be fully selfish rather than partially and selectively altruistic.

A specific application of the nonmonotonicity of altruism due to collective action effects is in the structure of collective wage bargaining institutions (Calmfors & Driffill, 1988). If unions are able to overcome the collective action problem at the industry level, this can result in higher wages and higher unemployment than a competitive labour market, or where collective bargaining occurs at the firm level. This is because the labour demand curve faced by a union is less elastic at the industry level, because consumers cannot go elsewhere for substitutes. If bargaining occurs at an economy-wide level through an incomes policy, on the other hand, the central union knows that it will only push up prices if it pushes up all wages, and so will exercise more restraint in nominal wage demands.

Moral hazard between altruist and beneficiary is another cause of normative limitation to altruism. This kind of effect relies upon some existing asymmetry in the altruism level of individuals, since if all individuals were fully altruistic they would not be tempted to do anything harmful to others. Stark and Bernheim present a version of the Samaritan's Dilemma (Bernheim & Stark, 1988). Take a three-stage game. Agents A and B each start off with a certain level of resources. First agent A decides how much to consume in period 1. Agent B then chooses his period 1 consumption. Agent A then chooses her period 2 consumption, transferring her residual resources to agent B. Suppose also that agent A is partially altruistic towards B whereas agent B is fully selfish. Agent B will have an incentive to consume an inefficiently high amount of his resources at stage 2, since he knows that agent A will compensate him for this at stage 3. In order to reduce agent B's incentive to do this, agent A will then consume an inefficiently large amount of her wealth at stage 1. This can lead to an outcome where higher altruism on the part of agent A results in a less efficient outcome.

1.18 *Consequences*

A central argument of this survey is that it would be desirable for variations in the level of altruism exhibited by agents to be a standard feature in analytic modelling, in a similar manner to imperfect information. This may potentially have far-reaching implications. Firstly, the impact of the presence of altruism has been analysed in the context of the theory of cost-benefit analysis. It has been shown that

in the presence of non-paternalistic altruism, household willingness to pay for a public good exceeds the sum of individuals' willingness to pay (Quiggin, 1998). An attempt to estimate the results of the presence of paternalistic altruism on the value of statistical life by calibrating to 2.5 individuals per altruistic family predicted that this value is 10%-40% higher than the individual value (Jones-Lee, 1992).

A number of overlapping generations models of environmental degradation have shown that the presence of partial altruism does not guarantee an efficient internalization of these externalities (Jouvet et al., 2000) (Turner, 1997). It has also been found that co-operation between nations to internalise current environmental externalities may lead to a deterioration of future environment relative to non-co-operation because improved environmental technology frees up more resources for consumption (John & Pecchenino, 1997). The first-best solution requires internalization of the inter- as well as intra-generational externalities via a specific optimal form of altruism.

The final area where the normative impact of altruism has been considered is in the literature on optimal taxation. It has been shown that altruism can either increase or decrease optimal Pigovian taxation depending on its precise form (Johansson, 1997). The optimal subsidy for voluntary giving has also been found to depend in a dramatic way upon the motivation for giving, particularly whether it is of the warm-glow or pure altruism kind (Kaplow, 1998).

Altruism has interesting consequences for many areas of economic theory. It should have a regular status alongside the self-interest assumption. However, the prevalence of altruism in human behaviour, along with its manifest imperfections, raises the much more fundamental question of whether the level of altruism in human society as a whole is suboptimal or superoptimal.

1.19 *The Sequential Punishment Model*

> [I]n comparison with a situation wherein altruism is absent altogether, the prevalence of just some altruism could result in Pareto *inferior* outcomes. Hence, if the formation of altruism may not only fail to do any good but may actually make things worse whereas the formation of sufficiently high levels of altruism is almost always beneficial,...a troubling discontinuity arises: to the extent that the formation of altruism is like the rising of bread dough (i.e. it *has* to be gradual) groups yearning to build up their social stock of altruism may have to endure Paretial *deterioration* before experiencing Paretial gains. Perhaps one reason why a great many societies consist of self-interested economic men and women rather than altruistic economic men and women has to do with this nonmonotonicity.
>
> (Stark, 1989)

These are stimulating conjectures from Stark and Bernheim. "The Socially Optimal Level of Altruism" (Chapter 2) corroborates many of them. It makes a contribution to the existing literature on the "limits to altruism" by using an appealingly abstract and general model of social interaction to illustrate that altruistic preferences on the part of individuals can often be unnecessary, and even counterproductive, from a social welfare perspective. More altruism is not always better for society. The most valuable additional insights offered are, on the one hand, that even the very high levels of altruism exhibited by individuals who care almost as much about others as themselves can sometimes be insufficient to "get the dough to rise" but that, on the other, it can be the case that even malevolent individuals who wish to harm one another could be induced to behave socially beneficially with the appropriate system of social control, and that this need not be more difficult than with altruistic individuals. Central to this result is the fact that altruism has multiple opposing effects upon incentives.

In the sequential punishment model, individuals receive opportunities to harm one another, one after another. If they inflict harm, they gain a random amount of felicity (whose value is known in advance), and the person they choose to hurt loses one unit of felicity. Individuals have free choice, and are indifferent, as to whom they punish. In a single-move game, individuals clearly must be sufficiently altruistic in order to prevent them from giving in to the temptation to harm others for their own benefit; altruism is always a good thing, because it reduces the temptation of individuals to misbehave (this is therefore referred to as the "temptation effect"). When there are a number of sequential moves, however, individuals can, contingent upon earlier observed misbehaviour, punish selfish miscreants. This creates the possibility of counterproductive altruism, both because more altruistic individuals are less willing to punish (the "willingness effect") and also because they are harder to hurt by threatening them in place of others (the "severity effect").

The sequential punishment model has relevance to issues which have arisen in recent empirical work on altruistic behaviour. The use of altruistic punishment has come to be seen by a number of empirical economists, influenced by a cultural evolution perspective, as the key to explaining how imperfect individual altruism can be "magnified" to achieve socially efficient outcomes (Fehr & Gächter, 2002a). Altruistic punishment is thus central to the operation of the social structure as an altruism amplification device. The presence of some individuals who are willing to punish others, even at harm to themselves, is at least as vital to achieving social efficiency as the presence of "traditional" altruists.

Players are assumed to be indifferent as to which other player they punish. This may seem an objectionably strong assumption, but it is a reasonable simplifying one since it is tantamount to assuming that preferences over who is punished are sufficiently small to enable punishment schemes to be incentivized. The idea that society may have punitive actions which can be cheaply directed onto miscreants is also an important one in experimental economics and cultural evolution theory. For example, Fehr et al. (Fehr & Gächter, 2002b) argue that altruistic punishment, the willingness to harm others at cost to oneself, is vital in making altruism work as the glue of human society, because it is usually less costly to harm others than to help them, and this enables cheaters to be credibly threatened with punishment. Sober and Wilson (Sober & Wilson, 1999) similarly argue that an important reason why human societies exhibit complex functional integration is the low cost of punishing transgressors against social norms (for example, by simply refusing to interact with them).

The complete solution to the model uses optimal punishment paths, as developed by Abreu for infinitely-repeated stage games (Abreu, 1986) (Abreu, 1988) (Lambson, 1987). The analysis of the sequential punishment model succeeds in showing that the severity and willingness effects sometimes outweigh the temptation effect; greater altruism can make it more difficult to achieve a socially efficient outcome. As the coefficient of altruism θ approaches 1 from below (i.e. as individuals become perfectly altruistic), the socially efficient equilibrium will always break down. Hence, to put the central point in a "nutshell", too much altruism is bad for society.

The introduction of some degree of preference by individuals over who they punish, via heterogeneous weightings applied to each other individual in the utility function, would greatly complicate the results of the model in that the conditions for co-operation with ongoing punishment would be different depending on the punisher and punishee. This would not, however, qualitatively change the property that punishment can be imposed to incentivize co-operation with a socially efficient initial path provided players are sufficiently patient. Equally importantly, although the precise result of the interaction between the temptation, willingness and severity effects would become more complex, their key role in driving the model would remain, and the result that too much altruism can sometimes be socially damaging would not be overturned.

Stark and Bernheim (Bernheim & Stark, 1988) have also considered this issue, and have established a theorem which shows that as altruism becomes perfect (by which we mean that equal weighting is

placed on the welfare of others relative to oneself), the Nash equilibrium outcome in a repeated game becomes arbitrarily close to the socially efficient outcome. Although it is strictly speaking a sequential rather than a repeated game, the sequential punishment model exhibits this property. However, it also exhibits the property that intermediate levels of altruism enable full social efficiency to be achieved, and that high enough levels of altruism which are still less than perfect cause the socially efficient equilibrium to break down, with a non-negligible negative effect on the efficiency of the equilibrium.

The sequential punishment model treats altruism in a general and abstract way, and provides good reasons to conclude that intermediate rather than high levels are the most socially desirable. Stark and Bernheim show an awareness of these possibilities in the following passage:

> While high levels of altruism are generally beneficial, this should provide little comfort to those who subscribe to the traditional views...The effects of altruism on economic interaction are complex, and must be assessed on a case-by-case basis.
>
> (Bernheim & Stark, 1988)

Stark and Bernheim discuss one of the factors which is of central concern in the sequential punishment model, in the form of the willingness effect. This they describe as the possibility that a more altruistic individual will be perceived as a "softy". Stark and Bernheim show, using a simple infinitely-repeated co-operation game, that altruism can, for this reason, make an efficient co-operative outcome harder to achieve. They also hint that a full analysis of optimal punishment could provide richer results in this area:

> In analysis of repeated games, it is standard practice to enforce co-operative outcomes through "Nash reversion"...While more severe punishments are frequently available..., Nash reversion is by far the simplest and most widely discussed method of enforcing co-operation. We employ it here.
>
> (Bernheim & Stark, 1988)

The sequential punishment model, although not a true repeated game, turns out both to be simple enough, and to share enough features with repeated games, to enable a satisfyingly full characterization of the optimal punishment paths, and thus a complete analysis of the effect of altruism on the sustainability of socially efficient equilibria. The most important general result from the model is that high levels of altruism close to, but not quite reaching, perfect utilitarian altruism unambiguously break the supportability constraint on the socially efficient outcome.

The sequential punishment model is intended to capture a general feature of the economic world in a simple but abstract manner. Other models with a similar idiom include Samuelson's "pension game" (Samuelson, 1958) and Diamond's model of fiat money in a "coconut economy" (Diamond, 1984). In Samuelson's model, an infinite series of individuals, each of whom lives for two periods, must decide whether to make a gift to the individual who is old when they are young. Young individuals have an endowment of one "chocolate", but the old have no endowment. Multiple subgame-perfect Nash equilibria exist, some where chocolate pensions are voluntarily given (because it is believed that receiving a pension from the next generation will be conditioned on having donated to the previous generation) and some where the pension system never "gets off the ground". This model captures, in a simple and elegant manner, the key features that make pension systems (either state or private) fragile, since they depend on the belief that the next generation will provide.

Diamond's model captures an essential point of economic life relevant to the microfoundations of macroeconomics. Production is only profitable when trading partners can be found. This creates the potential for multiple equilibria, with high and low levels of economic activity. The economy becomes a giant co-ordination game. The "parable" for the Diamond model involves a society of individuals who live on an island consisting of palm trees of varying height, each with a coconut at the top. Climbing trees is costly. Individuals can only gain utility from consuming their coconut by finding another individual to "swap" with. Each individual can only carry one coconut at a time. Whether or not it is worth climbing a particular tree therefore depends upon how many potential trading partners there are "out there". This simple abstract framework, and neat intuitive back-story, captures perfectly the existence of multiple search equilibria in a macroeconomic system which uses fiat money.

1.19.1 The evolutionary sequential punishment model

In a second paper, "Punishment and the Potency of Group Selection" (Chapter 3), a different aspect of the relationship between altruistic preferences and punishment systems is explored, and a new twist to the story offered. A simplified three-player two-move finite version of the sequential punishment model is used, but the preferences of the individuals playing the game are now permitted to evolve over time. It is shown that the use of punishment equilibria weakens the potency of the group selection mechanism, making it harder for altruistic preferences to evolve.

The intuition for this result is that group selection depends upon more altruistic groups doing better on average than more selfish groups. However, by making selfish individuals behave better, the use of a punishment system weakens this performance differential, thus weakening group-level selection pressures. The normative consequences of this phenomenon are ambiguous because there are two counteracting effects. The static effect improves social welfare, because, given a certain population composition, the use of punishment makes people behave better. The dynamic effect, that fewer altruists are able to evolve, can sometimes, however, outweigh the static gain, so that society is actually worse off in the long run when using a punishment system.

1.20 *Conclusion*

We have examined disparate evidence from a number of areas of economics in order to establish the pervasive presence of altruism in human motivation, but also the prevalence of imperfections. This survey has explored a number of positive and normative theoretical reasons for this phenomenon. In particular, it has presented a broader context for the innovations in the fields of economic and cultural evolution theory offered respectively in "The Socially Optimal Level of Altruism" (Chapter 2) and "Punishment and the Potency of Group Selection" (Chapter 3). The first of these papers shows that too high a level of altruism has a detrimental impact on the effectiveness of punishment systems to act as an "altruism amplification device". The second paper explores the reverse phenomenon - the use of a punishment system may weaken the ability of altruistic preferences to survive in a dynamic environment of cultural evolution.

References

Abel, Andrew B. and Warshawsky, Mark (1988). **"Specification of the Joy of Giving: Insights from Altruism"**. *The Review of Economics and Statistics*, 70(1), 145–149.

Abreu, Dilip (1986). **"Extremal Equilibria of Oligopolistic Supergames"**. *Journal of Economic Theory*, 39, 191–225.

Abreu, Dilip (1988). **"On the Theory of Infinitely Repeated Games with Discounting"**. *Econometrica*, 56(2), 383–396.

Adams, James D. (1980). **"Personal Wealth Transfers"**. *The Quarterly Journal of Economics*, 95(1), 159–179.

Agee, Mark D. and Crocker, Thomas D. (1996). **"Parents' Discount Rates for Child Quality"**. *Southern Economic Journal*, 63(1), 36–50.

Alesina, Alberto and Glaeser, Edward and Sacerdote, Bruce (2001). **"Why Doesn't the United States Have a European-Style Welfare State?"**. *Brookings Papers on Economic Activity*, 2001(2), 187–254.

Altig, David and Auerbach, Alan J. and Kotlikoff, Laurence J. and Smetters, Kent A. and Walliser, Jan (2001). **"Simulating Fundamental Tax Reform in the United States"**. *The American Economic Review*, 91(3), 574–595.

Altonji, Joseph G. and Hayashi, Fumio and Kotlikoff, Laurence J. (1992). **"Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data"**. *The American Economic Review*, 82(5), 1177–1198.

Altonji, Joseph G. and Hayashi, Fumio and Kotlikoff, Laurence J. (1997). **"Parental Altruism and Inter Vivos Transfers: Theory and Evidence"**. *The Journal of Political Economy*, 105(6), 1121–1166.

Andreoni, James (1989). **"Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence"**. *The Journal of Political Economy*, 97(6), 1447–1458.

Andreoni, James (1990). **"Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving"**. *The Economic Journal*, 100(401), 464–477.

Andreoni, James (1995a). **"Cooperation in Public-Goods Experiments: Kindness or Confusion?"**. *The American Economic Review*, 85(4), 891–904.

Andreoni, James (1995b). **"Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments"**. *The Quarterly Journal of Economics*, 110(1), 1–21.

Andreoni, James and Castillo, Marco and Petrie, Ragan (2003). **"What Do Bargainers' Preferences Look Like? Experiments with a Convex Ultimatum Game"**. *The American Economic Review*, 93(3), 672–685.

Andreoni, James and Miller, John (2002). **"Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism"**. *Econometrica*, 70(2), 737–753.

Andreoni, James and Miller, John H. (1993). **"Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence"**. *The Economic Journal*, 103(418), 570–585.

Arrow, Kenneth J. and Debreu, Gerard (1954). **"Existence of an Equilibrium for a Competitive Economy"**. *Econometrica*, 22(3), 265–290.

Barro, Robert J. (1974). **"Are Government Bonds Net Wealth?"**. *The Journal of Political Economy*, 82(6), 1095–1117.

Baumol, William J. (1975). **Business Responsibility and Economic Behaviour**. In E. S. Phelps (Ed.), *Altruism, Morality and Economic Theory* (pp. 45–57). Russell Sage Foundation, New York.

Becker, Gary S. (1974). **"A Theory of Social Interactions"**. *Journal of Political Economy*, 82, 1063–1093.

Becker, Gary S. (1981). **"Altruism in the Family and Selfishness in the Market Place"**. *Economica*, 48(189), 1–15.

Behrman, Jere R. and Pollak, Robert A. and Taubman, Paul (1986). **"Do Parents Favor Boys?"**. *International Economic Review*, 27(1), 33–54.

Behrman, Jere R. and Taubman, Paul (1986). **"Birth Order, Schooling, and Earnings"**. *Journal of Labor Economics*, 4(3), S121–S145.

Bergstrom, Theodore C. (1989). **"A Fresh Look at the Rotten Kid Theorem–and Other Household Mysteries"**. *The Journal of Political Economy*, 97(5), 1138–1159.

Bernheim, B. Douglas and Bagwell, Kyle (1988). **"Is Everything Neutral?"**. *The Journal of Political Economy*, 96(2), 308–338.

Bernheim, Douglas B. and Stark, Oded (1988). **"Altruism within the Family Reconsidered: Do Nice Guys Finish Last?"**. *The American Economic Review*, 78(5), 1034–1045.

Blackmore, Susan J. (1999). **The Meme Machine**. Oxford University Press, Oxford.

Boyd, R. and Richerson, Peter J. (1982). **"Cultural Transmission and the Evolution of Cooperative Behavior"**. *Human Ecology*, 10(3), 325–351.

Browning, Martin and Chiappori, Pierre-André (1998). **"Efficient Intra-Household Allocations: A General Characterization and Empirical Tests"**. *Econometrica*, 66(6), 1241–1278.

Bruce, Neil and Waldman, Michael (1990). **"The Rotten-Kid Theorem Meets the Samaritan's Dilemma"**. *The Quarterly Journal of Economics*, 105(1), 155–165.

Calmfors, Lars and Driffill, John (1988). **"Bargaining Structure, Corporatism and Macroeconomic Performance"**. *Economic Policy*, 6, 13–61.

Camerer, Colin and Thaler, Richard H. (1995). **"Anomalies: Ultimatums, Dictators and Manners"**. *The Journal of Economic Perspectives*, 9(2), 209–219.

Case, Anne and Lin, I. Fen and McLanahan, Sara (1999). **"Household Resource Allocation in Stepfamilies: Darwin Reflects on the Plight of Cinderella"**. *The American Economic Review*, 89(2), 234–238.

Chiappori, Pierre-André (1992). **"Collective Labor Supply and Welfare"**. *The Journal of Political Economy*, 100(3), 437–467.

Clark, Jeremy (1998). **"Fairness in Public Good Provision: An Investigation of Preferences for Equality and Proportionality"**. *The Canadian Journal of Economics*, 31(3), 708–729.

Cohen, Gerry A. (1978). **Karl Marx's Theory of History: A Defence**. Clarendon Press, Oxford.

Collard, David (1978). **Altruism and Economy**. Martin Robertson Ltd, Oxford.

Cox, Donald (1987). **"Motives for Private Income Transfers"**. *The Journal of Political Economy*, 95(3), 508–546.

Cox, Donald and Rank, Mark R. (1992). **"Inter-Vivos Transfers and Intergenerational Exchange"**. *The Review of Economics and Statistics*, 74(2), 305–314.

d'Aspremont, Claude and Gevers, Louis (1977). **"Equity and the Informational Basis of Collective Choice"**. *Review of Economic Studies*, 44(2), 199–209.

Davies, James B. and Zhang, Junsen (1995). **"Gender Bias, Investments in Children, and Bequests"**. *International Economic Review*, 36(3), 795–818.

Dawes, Robyn M. and Thaler, Richard H. (1988). **"Anomalies: Cooperation"**. *The Journal of Economic Perspectives*, 2(3), 187–197.

Dawkins, Richard (1976). **The Selfish Gene**. Oxford University Press, Oxford.

Diamond, Peter (1984). **"Money in Search Equilibrium"**. *Econometrica*, 52(1), 1–20.

Dowrick, Steve and Dunlop, Yvonne and Quiggin, John (1998). **"The Cost of Life Expectancy and the Implicit Social Valuation of Life"**. *The Scandinavian Journal of Economics*, 100(4), 673–691.

Fehr, Ernst and Fischbacher, Urs (2003). **"The Nature of Human Altruism"**. *Nature*, 425, 785–791.

Fehr, Ernst and Gächter, Simon (2000a). **"Cooperation and Punishment in Public Goods Experiments"**. *The American Economic Review*, 90(4), 980–994.

Fehr, Ernst and Gächter, Simon (2000b). **"Fairness and Retaliation: The Economics of Reciprocity"**. *The Journal of Economic Perspectives*, 14(3), 159–181.

Fehr, Ernst and Gächter, Simon (2002a). **"Altruistic Punishment in Humans"**. *Nature*, 415, 137–140.

Fehr, Ernst and Gächter, Simon (2002b). **"Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms"**. *Human Nature*, 13, 1–25.

Frank, Robert H. and Gilovich, Thomas D. and Regan, Dennis T. (1996). **"Do Economists Make Bad Citizens?"**. *The Journal of Economic Perspectives*, 10(1), 187–192.

Freeman, Katherine B. (1984). **"The Significance of Motivational Variables in International Public Welfare Expenditures"**. *Economic Development and Cultural Change*, 32(4), 725–748.

Freeman, Richard B. (1997). **"Working for Nothing: The Supply of Volunteer Labor"**. *Journal of Labor Economics*, 15(1), S140–S166.

Gevers, Louis (1979). **"On Interpersonal Comparability and Social Welfare Orderings"**. *Econometrica*, 47(1), 75–89.

Hammond, Peter (1975). **Charity: Altruism or Cooperative Egoism?** In E. S. Phelps (Ed.), *Altruism, Morality and Economic Theory* (pp. 115–131). Russell Sage Foundation, New York.

Hanemann, W. Michael (1994). **"Valuing the Environment Through Contingent Valuation"**. *The Journal of Economic Perspectives*, 8(4), 19–43.

Harsanyi, John C. (1986). **Utilitarian Morality in a World of Very Half-hearted Altruists**. In R. M. Heller, Walter P. Starr & D. A. Starrett (Eds.), *Social Choice and Public Decision Making* (pp. 57–73). Cambridge University Press.

Hayashi, Fumio (1995). **"Is the Japanese Extended Family Altruistically Linked? A Test Based on Engel Curves"**. *The Journal of Political Economy*, 103(3), 661–674.

Hayek, Friedrich A. (1960). **The Constitution of Liberty**. Routledge and Kegan Paul Ltd, London.

Heckathorn, Douglas D. (1990). **"Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control"**. *American Sociological Review*, 55(3), 366–384.

Hoffman, Elizabeth and Spitzer, Matthew L. (1985). **"Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice"**. *The Journal of Legal Studies*, 14(2), 259–297.

Hollander, Heinz (1990). **"A Social Exchange Approach to Voluntary Cooperation"**. *The American Economic Review*, 80(5), 1157–1167.

Holmes, Thomas P. (1990). **"Self-Interest, Altruism, and Health-Risk Reduction: An Economic Analysis of Voting Behavior"**. *Land Economics*, 66(2), 140–149.

Johansson, Olof (1997). **"Optimal Pigovian Taxes under Altruism"**. *Land Economics*, 73(3), 297–308.

John, A. Andrew and Pecchenino, Rowena A. (1997). **"International and Intergenerational Environmental Externalities"**. *The Scandinavian Journal of Economics*, 99(3), 371–387.

Jones-Lee, Michael. W. (1992). **"Paternalistic Altruism and the Value of Statistical Life"**. *The Economic Journal*, 102(410), 80–90.

Jouvet, Pierre-Andre and Michel, Philippe and Vidal, Jean-Pierre (2000). **"Intergenerational Altruism and the Environment"**. *The Scandinavian Journal of Economics*, 102(1), 135–150.

Kaplow, Louis (1998). **"Tax Policy and Gifts"**. *The American Economic Review*, 88(2), 283–288.

Kotlikoff, Laurence J. (1988). **"Intergenerational Transfers and Savings"**. *The Journal of Economic Perspectives*, 2(2), 41–58.

Kotlikoff, Laurence J. and Spivak, Avia (1981). **"The Family as an Incomplete Annuities Market"**. *The Journal of Political Economy*, 89(2), 372–391.

Kuehlwein, Michael (1993). **"Life-Cycle and Altruistic Theories of Saving with Lifetime Uncertainty"**. *The Review of Economics and Statistics*, 75(1), 38–47.

Kuhn, Thomas S. (1970). **The Structure of Scientific Revolutions**. University of Chicago Press, Chicago.

Kuhn, Thomas S. (1977). **Objectivity, Value Judgement and Theory Choice**. In *The Essential Tension*. University of Chicago Press, Chicago.

Kurz, Mordecai (1984). **"Capital Accumulation and the Characteristics of Private Intergenerational Transfers"**. *Economica*, 51(201), 1–22.

Laitner, John and Juster, F. Thomas (1996). **"New Evidence on Altruism: A Study of TIAA-CREF Retirees"**. *The American Economic Review*, 86(4), 893–908.

Lakatos, Imre (1970). **Falsification and the Methodology of Social Science Research Programmes**. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*. Cambridge University Press, Cambridge.

Lambson, Val Eugene (1987). **"Optimal Penal Codes in Price-Setting Supergames with Capacity Constraints"**. *Review of Economic Studies*, 54(3), 385–97.

Lucas, Robert Jr (1976). **"Econometric Policy Evaluation: A Critique"**. *Carnegie-Rochester Conference Series on Public Policy*, 1(1), 19–46.

Mandeville, B. (1732). **The Fable of the Bees or Private Vices, Publick Benefits**. Liberty Fund, Indianapolis.

McGranahan, Leslie Moscow (2000). **"Charity and the Bequest Motive: Evidence from Seventeenth-Century Wills"**. *The Journal of Political Economy*, 108(6), 1270–1291.

McKean, Roland N. (1975). **Economics of Trust, Altruism and Corporate Responsibility**. In E. S. Phelps (Ed.), *Altruism, Morality and Economic Theory* (pp. 29–44). Russell Sage Foundation, New York.

McKelvey, Richard D. and Palfrey, Thomas R. (1992). **"An Experimental Study of the Centipede Game"**. *Econometrica*, 60(4), 803–836.

Meade, James E. (1973). **Theory of Economic Externalities: The Control of Environmental Pollution and Similar Social Costs**. Sijthoff, Leiden.

Merton, Robert K. (1968). **Social Theory and Social Structure**. Free Press, New York.

Olson, Mancur (1965). **The Logic of Collective Action**. Harvard Economic Studies, Harvard.

Pattanaik, Prasanta K. (1971). **Voting and Collective Choice**. Cambridge University Press, Cambridge.

Popp, David (2001). **"Altruism and the Demand for Environmental Quality"**. *Land Economics*, 77(3), 339–349.

Popper, Karl (1959). **The Logic of Scientific Discovery**. Hutchinson Press, London.

Quiggin, John (1998). **"Individual and Household Willingness to Pay for Public Goods"**. *American Journal of Agricultural Economics*, 80(1), 58–63.

Rangazas, Peter C. (1996). **"Fiscal Policy and Endogenous Growth in a Bequest-Constrained Economy"**. *Oxford Economic Papers*, 48(1), 52–74.

Roberts, Kevin W S (1980). **"Interpersonal Comparability and Social Choice Theory"**. *Review of Economic Studies*, 47(2), 421–39.

ROTEMBERG, JULIO J. (1994). **"Human Relations in the Workplace"**. *The Journal of Political Economy*, 102(4), 684–717.

SAMUELSON, PAUL A. (1958). **"An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money"**. *Journal of Political Economy*, 66, 467.

SEATER, JOHN J. (1993). **"Ricardian Equivalence"**. *Journal of Economic Literature*, 31(1), 142–190.

SEN, AMARTYA (1974). **"Informational bases of alternative welfare approaches : Aggregation and income distribution"**. *Journal of Public Economics*, 3(4), 387–403.

SIMON, HERBERT A. (1993). **"Altruism and Economics"**. *The American Economic Review*, 83(2), 156–161.

SLOAN, FRANK A. AND ZHANG, HAROLD H. AND WANG, JINGSHU (2002). **"Upstream Intergenerational Transfers"**. *Southern Economic Journal*, 69(2), 363–380.

SMITH, ADAM [1776] (1976). **An Inquiry into the Nature and Causes of the Wealth of Nations**. Oxford University Press, Oxford.

SOBER, ELLIOT AND WILSON, DAVID S. (1999). **Unto Others**. Harvard University Press, Cambridge, Massachussets.

SOLTIS, JOSEPH AND BOYD, ROBERT AND RICHERSON, PETER J. (1995). **"Can Group-Functional Behaviors Evolve by Cultural Group Selection?: An Empirical Test"**. *Current Anthropology*, 36(3), 473–494.

STARK, ODED (1989). **"Altruism and the Quality of Life"**. *The American Economic Review*, 79(2), 86–90.

STARK, ODED (1995). **Altruism and Beyond**. Cambridge University Press, Cambridge.

STARK, ODED AND FALK, ITA (1998). **"Transfers, Empathy Formation, and Reverse Transfers"**. *The American Economic Review*, 88(2), 271–276.

Stevens, Thomas H. and More, Thomas A. and Glass, Ronald J. (1994). **"Interpretation and Temporal Stability of CV Bids for Wildlife Existence: A Panel Study"**. *Land Economics*, 70(3), 355–363.

Sugden, Robert (1982). **"On the Economics of Philanthropy"**. *The Economic Journal*, 92(366), 341–350.

Sugden, Robert (1984). **"Reciprocity: The Supply of Public Goods Through Voluntary Contributions"**. *The Economic Journal*, 94(376), 772–787.

Tcha, Moonjoong (1996). **"Altruism and Migration: Evidence from Korea and the United States"**. *Economic Development and Cultural Change*, 44(4), 859–878.

Turner, Matthew A. (1997). **"Parental Altruism and Common Property Regulation"**. *The Canadian Journal of Economics*, 30(4a), 809–821.

Weaver, Robert D. (1996). **"Prosocial Behavior: Private Contributions to Agriculture's Impact on the Environment"**. *Land Economics*, 72(2), 231–247.

Weiss, Yoram and Willis, Robert J. (1993). **"Transfers among Divorced Couples: Evidence and Interpretation"**. *Journal of Labor Economics*, 11(4), 629–679.

Wilhelm, Mark O. (1996). **"Bequest Behavior and the Effect of Heirs' Earnings: Testing the Altruistic Model of Bequests"**. *The American Economic Review*, 86(4), 874–892.

Willis, Robert J. (1999). **"A Theory of Out-of-Wedlock Childbearing"**. *The Journal of Political Economy*, 107(6), S33–S64.