Soc Choice Welfare (1985) 2:221–234

Social Choice and Welfare

The Welfare Economics of the Future

A Review of *Reasons and Persons* by Derek Parfit*

John Broome

Department of Economics, Alfred Marshall Building, 40 Berkely Square, Bristol BS8 1HY, England

Accepted June 19, 1985

I. Self-Interest in Economics

The ordinary theory of behaviour in economics does not assume that people act in their own interest. It takes people to maximize their own utility, but their utility is defined as what they maximize. More precisely, one alternative is said to have more utility for a person than another if, were the person to have a choice between the two, she would choose the former[1]. I shall be using 'utility' in this sense throughout this review. A person's utility, then, need not represent her interest. All the theory requires is enough consistency in a person's choices to allow a utility function to be defined.

But welfare economics, as it is generally interpreted, does assume that people act in their own interest. It assumes that increasing a person's utility is in her interest. But, like the theory of behaviour, it takes a person's utility to represent her choices. It assumes, therefore, that what a person would choose, given a choice of alternatives, coincides with what would be in her interest.

It is commonly supposed, however, that people sometimes discount their future good when making choices[2]. If this is so, then welfare economics is wrong to assume that people act in their own interest. If a person discounts, she will sometimes sacrifice a greater amount of future good for the sake of a smaller amount of present good. But a person is equally herself at every stage of her life, and good that comes to her at any stage is equally her good. This person is therefore choosing her own lesser good in preference to her own greater good. She is not acting in her own interest.

* I have benefited very greatly from discussion with members of the Economics and Philosophy Departments of Bristol University about *Reasons and Persons* and about this review. And I am particularly grateful to Deborah Mabbet, Adam Morton and Amartya Sen for their comments

[1] This is not quite right. If a person is indifferent between two alternatives, it will still be true that were she to have a choice between them, she would choose one or the other. Yet their utilities must be equal. Indifference makes it difficult to define utility in terms of choice, but I shall ignore this difficulty

[2] Pigou, for one, believed this (1932, pp. 24–5)

Suppose a government could adopt some policy that would raise everbody's utility. Each person, that is to say, would choose to have this policy adopted if it was up to her. Let us call it 'the popular policy'. Perhaps it increases present consumption at the expense of investment for the future. Ordinary welfare economics would favour this policy on the grounds that it would be good for everybody. But if people discount, we have just seen that actually it might not be good for everybody.

What should be done about this? We might take a conservative line and continue to support the conclusions of ordinary welfare economics; we might continue to think it right to increase a person's utility, provided it does not reduce someone else's. So we would continue to favour the popular policy. Alternatively we might take a revisionist line and draw different conclusions. A basis for revisionism can be found in Derek Parfit's *Reasons and Persons* (1984). But first I want to consider what basis can be found for conservatism.

II. Discounting Future Good

One source of conservatism is a sceptical attitude that is common in economics towards the notion of good. Let us look more closely at what it means to say a person discounts her future good.

Suppose a person's life has T periods, and let $c_t = (c_{t1}, c_{t2}, \ldots, c_{tn})$ be the vector of her consumption in the t'th. Represent the choices she would make at t by a utility function $U_t(c_1, c_2, \ldots, c_T)$. So $U_t(c) > U_t(\bar{c})$ implies that if the person were to have a choice at t between the consumption sequences c and $\bar{c}$ she would choose c. In one sense, this person discounts if at the margin she is willing to give up more than one unit of some commodity in the further future for the sake of getting an extra unit of it in the nearer future. That is: if

$$\frac{\partial U_t/\partial c_{\tau' r}}{\partial U_t/\partial c_{\tau r}} < 1 \qquad \text{for some } r \text{ and some } \tau,\ \tau' \text{ with } t \leqq \tau < \tau' \leqq T.$$

But this is not what it means to say the person discounts her future good. At the margin she may simply have more use for this commodity in the nearer future.

To do better we need a notion of the person's good, or her interest, or how well her life goes. Her good will depend (partly) on her consumption. Write it $V(c_1, c_2, \ldots, c_T)$. A better shot at a definition is to say this person discounts her future good if the marginal rate of substitution mentioned above is less than is in her interest:

$$\frac{\partial U_t/\partial c_{\tau' r}}{\partial U_t/\partial c_{\tau r}} < \frac{\partial V/\partial c_{\tau' r}}{\partial V/\partial c_{\tau r}} \qquad \text{for some } r \text{ and some } \tau,\ \tau' \text{ with } t \leqq \tau < \leqq \tau' \leqq T.$$

But this is not quite right either. The benefit from a commodity does not always

come at the time when it is consumed[3]. And so a person may overindulge in a present commodity because she is concerned too much for her future good, rather than too little. A student who overindulges in midnight oil is an example.

To define discounting accurately we need the further notion of the person's good at a particular time t. Write it v_t. For generality we should treat v_t as a function $v_t(c_1, c_2, \ldots, c_T)$ of her consumption in every period. The person's total good will depend, partly at least[4], on her good in each period. So we may write (shifting the notation):

$$V = V(v_1, v_2, \ldots, v_T).$$

Apart from discounting, there are other ways in which a person's choices may fail to achieve her own good. But let us abstract from these. So let us suppose that our subject chooses amongst consumption sequences only on the basis of the good they bring her in the various periods of her life. Then

$$U_t = U_t(v_1, v_2, \ldots, v_T).$$

Now, when I say that the person discounts her future good, I mean that

$$\frac{\partial U_t/\partial v_{\tau'}}{\partial U_t/\partial v_\tau} < \frac{\partial V/\partial v_{\tau'}}{\partial V/\partial v_\tau} \quad \text{for some } \tau \text{ and } \tau' \text{ with } t \leqq \tau < \tau' \leqq T.$$

So even to make sense of the claim that a person discounts her future good, we need a notion of a person's good and also a notion of her good at a time, and these notions must be independent of the choices she would make at any time. But many economists are sceptical about notions of good that are severed from choices. They think they make no sense. And this will lead them to object to revisionism in welfare economics. It does not really make sense, the objection goes, to say a person discounts her future good. So discounting can give us no reason to revise the conclusions of welfare economics.

III. Conservatism

Parfit (like every other moral philosopher I can think of) is happy to use such notions of good. So a truly sceptical economist will be out of sympathy with his arguments. I suspect, myself, that this sort of scepticism is generally an affectation. I think we all understand very well what it is for a person to act imprudently against her own interest. And we meet actual cases of it daily. But in any case scepticism itself leaves welfare economics up in the air. It is not itself enough to justify

---

[3] A sophisticated definition of 'consumption' might make consumption coincide with the benefit it brings. But to set up such a definition would require the notion of the person's good at a particular time, which is described in the next paragraph

[4] There are several important and interesting questions here. Does all the good that comes to a person have to be datable to a particular period of her life, or are there undated goods? Can the goodness of a period be measured cardinally? Is intertemporal comparison possible between goods in different periods? Should V be simply the sum of the v's? Unfortunately, I shall have to leave these questions aside

conservatism with. Conservatism claims that, other things being equal, it is right to make a change that raises somebody's utility. To justify this, either it needs to be shown that raising a person's utility is necessarily good for her, as welfare economics has traditionally assumed, or else some other grounds must be found for it.

To claim that raising a person's utility is necessarily good for her is, I think, hopeless. The trouble is that a person has many utility functions, one for each time in her life. Normally they will order her consumptions streams differently, and none of them has any better claim than the others to represent her good. Sometimes it may happen that the utility functions are 'dynamically consistent': whenever $1 \leqq t < t' \leqq T$, U_t and $U_{t'}$ agree in the ordering of sequences $c_{t'}, \ldots, c_T$, given $c_1, \ldots, c_{t'-1}$. (In simple cases this will happen if the person discounts exponentially.) If a person's utility functions are dynamically consistent, she can at any time – the present for instance – plan her future consumption and at no later time will she want to change her plan. At a later time she can only make plans about what is then the future, and there her preferences will coincide with what they are now. It may look as though this gives her present utility function a special claim to represent her good in so far as it is affected by present and future consumption. Nothing can now be done about her past consumption. But if we choose her present and future consumptions so as to maximize her present utility, are we not then, given her past consumption, maximizing her good? The answer is no. Suppose t is now and t' sometime in the future. We are assuming that U_t and $U_{t'}$ coincide in their ordering of consumption beyond t'. But they will not normally coincide in their ordering of consumptions between t and t'. At t' the person will regret some of the choices she made at t. That is to say, were she *per impossible* to have those choices again, she would make them differently[5]. If she starts smoking at t, by t' she may wish she had not, even though at t' she may continue smoking as many cigarettes as, at t, she planned to smoke. If so, there are no grounds for saying that it is in her interest to start smoking at t. Even dynamic consistency then, does not give one utility function rather than another any special claim to represent a person's interest. If a person has different utility functions at different times in her life – and she normally will – it cannot plausibly be claimed that any one of them represents her good.

Conservatism, then, will have to find some other reason why it should be right to raise a person's utility. The obvious one is that that is what she would want. We might give, as a reason for adopting the popular policy in my example, simply that everyone wants it.

Why should it be right to adopt a policy that everyone wants? We cannot now answer this by saying that it is good for everyone. There is to be sure, a respectable theory of value that says it is good for a person to get what she wants. But I hope we have agreed by now that the popular policy may not be good for everyone. Everyone wants it now. But later they may think differently. The theory that it is good for a person to get what she wants must take account, not just of what she wants now, but of what she wants at every stage of her life.

---

[5] In this sense, if a person discounts she will nearly always regret her past decisions. Parfit (1984, pp. 187-8) is either wrong about this or using 'regret' in a different sense

The argument has to be that the popular policy should be adopted even though it may not be good for everyone, or indeed for anyone. The reason for favouring the policy cannot, therefore, be that it is beneficial[6]. Conservatism will have to see the purpose of welfare economics as something other than beneficence. But this requires a very profound change in the spirit of welfare economics. Traditionally welfare economics has always been concerned with finding policies that do the most good. If, now, it is to recommend policies that do not do good, and may do harm, I do not see how it could even decently continue to be called 'welfare economics'. Conservatism turns out to be very radical indeed.

Of course there *are* things to be said in favour of doing what everyone wants. One is that it is democratic. Democracy is a reason for favouring the popular policy. It favours it, not because there is something good about its results, but because there is something good about the process for arriving at it. Basing welfare economics on reasons like this absorbs it into politics. Usually people's wants conflict. When they do, welfare economics traditionally looks for the best way of weighing one person's good against another's. But one might look instead for a mechanism for resolving the conflict, and the mechanism need not necessarily be judged by the good effects of the policies it leads to. We may simply want a fair, democratic mechanism.

Social choice theory may be interpreted this way. The standard interpretation is that, weighing up people's conflicting preferences, it looks for the best 'social choice'. But it might also be interpreted as looking for a satisfactory process for resolving the conflict. The difference is that under this second interpretation there is no need to claim that the result of the process is beneficial. Merit is lodged in the process itself. And the rest of welfare economics could be given a similar political interpretation. I do not think that many if its conclusions could survive the reinterpretation unaltered. Social choice theorists, for instance, generally impose consistency conditions – transitivity and the like – on social choice, and it is hard to see why these should be required if they were only looking for a good process rather than good results[7]. Nevertheless, this seems to be the way the conservative line is forced to go.

IV. Is Discounting Rational?

Now let us turn to the revisionist alternative. The obvious revision to propose is to base welfare economics on people's interests, rather than on what they would choose. If the popular policy turns out to be bad for most people, and is popular only because of discounting, according to revisionism this policy should be rejected. Can this be justified?

---

[6] One might say that, though not good for everyone, it is good for everyone now. This might make it beneficial in a sense. Perhaps the popular policy is indeed better for everyone now than the alternative. But, speaking generally, increasing a person's present utility is not necessarily good for her now. In the notation of Sect. 2 a person's present good is given by v_t, not U_t. And if the person takes any thought at all for the future v_t and U_t will be different

[7] Compare Sen (forthcoming)

All would be straightforward, perhaps, if we could say that people are irrational when they discount. Irrationality in choice is perhaps something the government should correct for. (I shall be saying something later about the paternalism implicit in this.) A.C. Pigou thought that discounting is irrational (1932, p. 25) and he thought that 'the State should protect the interests of the future in some degree against the effects of our irrational discounting'. (p. 29) But if, on the other hand, discounting is not irrational then revisionism might seem harder to justify. How could a government be justified in overriding people's own choices if these choices are perfectly rational?

So we must ask whether discounting is irrational. At this point we shall have to begin following closely Derek Parfit's arguments in *Reasons and Persons*. For many years to come this book will be the recognized starting point for discussions of moral attitudes to the future. We need it now because Parfit considers very thoroughly whether discounting is necessarily irrational. His conclusion is that it is not.

There is no doubt that discounting is against a person's own interest. Does this not immediately imply that it is irrational? It does, of course, if we assume that rationality requires a person always to act in her own interest. This assumption Parfit calls 'the self-interest theory'. One of the principal aims of his book is to refute it.

Parfit's argument against the self-interest theory occupies Part II of *Reasons and Persons*. Against it he sets up two alternatives. One is 'the present-aim theory'. This says that it is not irrational for a person now to pursue the aims she now has, even if it is against her own interest. So if she now cares more for her present good than for her future good it is not irrational for her to sacrifice more of her future good for the sake of less of her present good. (This needs qualification – I shall come to that.) Also against the self-interest theory Parfit sets up a moral view: it is not irrational for a person to act in other people's interests, even if it is against her own.

These two theories oppose the self-interest theory from opposite directions. To combat them the self-interest theory has, Parfit says, to fight a war on two fronts. The question is how rationality requires me, now, to act. Against morality the self-interest theory has to argue that there is something special about good that comes to *me*, so that I, now, should act to promote this good only. Against the present-aim theory it has to argue that there is nothing special about the aims I have *now*, so that I, now, should not act to promote those aims only. But there is an analogy between 'me' and 'now'. The meaning of both is relative to the occasion when they are uttered: 'me' refers to the utterer, 'now' to the time of utterance. The self-interest theory is partially relative. It attaches importance to me but not to now. Arguments it can use against morality, showing something special about me, tend to stress the importance of relativity. So they tend to show something special about now too. And arguments it can use against the present-aim theory, showing there is nothing special about now, tend to stress the importance of non-relativity. So they tend to show there is nothing special about me either. Parfit works out this idea in detail. He concludes that the self-interest theory's position, half way between the fully relative present-aim theory and the non-relative morality, is untenable.

This, of course, does not commit him to either of the two alternatives. Furthermore, so far as the present-aim theory is concerned he favours a 'critical' version of it. This version does not accept all of a person's present aims uncritically;

some may be irrational. It may indeed even rule out as irrational the aim of promoting one's present good at the expense of a greater loss of future good. Parfit allows this as a possibility. At this stage, then, he still allows as a possibility that discounting may turn out to be irrational. So far, he has only knocked out one argument to the effect that it definitely must be. Parfit's positive argument that discounting can actually be rational occurs in Part III.

V. Personal Identity

Part III of *Reasons and Persons* is about personal identity. It is the book's most polished section. It presents arguments that Parfit has worked out in a series of articles over more than ten years. His views on this subject seem settled. Here his style is at its most powerful. He writes in short, lucid sentences. He uses a multitude of strange and disturbing examples. He piles complex argument on complex argument. He beats down opposition, one might almost say, by sheer weight of argument. One is left feeling rather numb. And what he is arguing for with such power is a strikingly original viewpoint that can alter one's whole attitude to life, to other people and to death. According to his own testimony (pp. 281-2), it has done so for him. Part III is a *tour de force*. Elsewhere in the book the style is the same. The sentences are lucid, but the arguments are sometimes so complex that it takes patience to find one's way through. Parfit is always anxious to make a proper reply to every opinion he disagrees with. (An exception is the peremptory dismissal on p. 337 of the view that interpersonal comparisons of harm are senseless: "I shall here ignore this view". Just what it deserves.) Consequently the flow of the book is interrupted by many detours, and several levels of argument are often nested within each other. Not surprisingly, this seems to happen most often in areas where Parfit's own opinions are still developing. But the effort required to unravel the argument is always rewarded. Arguments as careful and thorough as this, which block every avenue of dissent, can scarcely fail to be convincing in the end. No one, I think, could read this book without being forced to change her opinion about something.

You hear on the news that next week's prize in the national lottery will be a big one. You wonder whether the winner will be you or someone else. Naturally this is a matter of interest to you. A prize coming to you has a different significance to you from a prize coming to someone else. Parfit's first task in Part III is to give an account of what it is for the person who next week gets the cheque in the post to be you, the very same person as is now listening to the radio and dreaming. His answer to this question about personal identity is reductionist. If a person identified at one time (the prizewinner, say) is the same as a person identified at another time (the person listening to the news), Parfit believes this fact can be reduced to some other facts that can be described without using the concept of personal identity. The facts he particularly has in mind are the existence of certain psychological links between the person identified at one time and the person identified at the other. If it is you who wins the prize then the prizewinner will remember doing some of the things you do now. And she will do some of the things you now intend to do. These links of intention and memory, and other psychological links, are, according to Parfit, part of what goes to make it true that she and you are the same person.

But giving an account of personal identity is not Parfit's primary purpose. He gives one in order to clear the way for what really is his primary purpose. He wants to argue that, when you hear about this prize, you are wrong to be particularly concerned about whether or not the winner will be you. What really matters to you is not the fact of identity itself – whether or not the winner will be you – but the existence of the psychological links that, partly, this fact reduces to. What matters is whether or not the winner will be psychologically linked to you in the special way that normally goes to make you identical. You have a special concern for what happens to you. But actually you should attach this concern to anyone who is psychologically linked to you in the appropriate way. Normally, of course, such a person will be you and no one else. Normally only you will remember what you do, only you will carry out your intentions, and so on. Wondering whether next week's prizewinner will be you comes in practice to the same thing as wondering whether she will be psychologically linked to you. It generally takes science-fiction examples like brain transplants and teletransportation (which Parfit uses liberally) to prise identity apart from the links. But there is one mundane way in which they come apart to some extent in everyday life.

Look ahead a few decades. The person you will then be will have rather weaker psychological links to you now than will the person you will be tomorrow. Memory fades. You will, of course, still be you in your old age. But according to Parfit what is significant to you now is not this fact of identity, but the psychological links. You now have a special concern for your own good. But this theory says that concern should really be attached, not to your own good exactly, but to the good of a person who is psychologically linked to you now. And since you in your old age will be only rather weakly linked psychologically to you now, it is reasonable for you now not to concern yourself much with the good you will then receive. So far as you now are concerned, the further in the future good comes to you, the less will it have the special importance you attach to your own good. It will have, for you now, more of the status of other people's good.

This, at last, is why Parfit thinks discounting of future good need not be irrational. It can be justified by the weakening of psychological links.

And now we can see what to do about discounting in welfare economics. That your distant future good is a bit like someone else's good may, perhaps, be a reason for you now to be less concerned about it. But it is not a reason for the government or a welfare economist to be less concerned about it. They should be equally concerned about everybody. It should not matter to them whether or not the person living in the future will be you, or how closely she will be psychologically connected to you now. If you do not now care so much about your future good, then your future self is under-represented in your present choices. Welfare economics will need to correct for that, just as it will for anyone else who is not fully represented in people's present choices, for anyone who is not yet born, for instance. Welfare economics, then, should concern itself with your good throughout your life, even if you yourself now do not. It should not discount your future good, even if you do.

This reductionist theory puts revisionism in a new light. Revisionism founds welfare economics on what is in people's interests rather than on what they would choose. This may seem intolerably paternalistic towards grown-up people. But according to Parfit the reason for doing so is to protect the interests of the people's

future selves, which are to be thought of as a bit like future people. The government, Pigou thought (1932, p. 29), is the trustee for future generations, and there is nothing unreasonably paternalistic about that. We could think of its trusteeship for future selves in a similar way (see Parfit p. 321).

VI. Distributive Justice

For a moment I want to step aside from our main line of enquiry. The reductionist theory of personal identity is important for another subject that has interested economists: distributive justice. Parfit considers this in Chap. 15.

When we look at the amounts of good that come to a person in the successive periods of her life the fact that they come to the same person is not, according to this theory, as significant as it may have seemed. What ties them together, and makes it morally significant that they are in the same life, is the psychological links between the periods of the person life. And, at least for widely separated periods, these links may be weak. Consequently, if reductionism is true, the moral principles that govern distribution between people should to some extent also apply within a single life. Suppose, for instance, that it is desirable for good to be equally distributed between people. Then to some extent it must also be desirable for good to be evenly distributed within a life. When assessing inequality in a society we should be less interested in comparing people's lifetime totals of good and more in comparing the good people enjoy at each stage in life. If everybody suffers a spartan youth but enjoys a golden old age, we should count the society as unequal.

Reductionism, then, suggests a new interpretation for the principle of equality. Parfit also believes it provides an argument – not by any means conclusive – against the principle itself. It suggests, he thinks, that we should not be so much concerned about equality and more concerned about the total of good, whoever it comes to. He thinks, then, that reductionism tends to favour utilitarianism. Now, I am not convinced by Parfit's own argument for this point (pp. 329–342). But I do think for other reasons that there is an affinity between reductionism and utilitarianism. The link is to be found in some theorems of economics that connect separability in utility functions to the additive form of utility functions (Gorman 1968). Reductionism seems to provide a reason for thinking that the goodness of a person's life should be separable into the goodness of its periods, and additive utility functions have a utilitarian tendency. (I can speak only in the vaguest terms here[8].) J.A. Mirrlees (1982) applies these theorems in defending utilitarianism, and in the course of doing so he shows a strong implicit commitment to reductionism. He argues, for instance, (p. 66) that if the psychological links between the periods of a person's life were completely severed, then the good she receives in the different periods should be separable. But Mirrlees's linking of reductionism and utilitarianism is neither explicit nor by any means complete. I believe, however, that a firm link can be made[9], and that it will contribute a great deal to our understanding of utilitarianism. This is perhaps one of the most fruitful areas where economics can contribute to moral philosophy.

---

[8] I have been more precise in Broome (1983)

[9] I have tried to make a start on the task (Broome 1983)

VII. Population

Let us return now to our main question. How is a person's future good to be treated in welfare economics, when the person herself discounts it? The suggestion that comes out of Parfit's work is to treat it in something like the way in which we would treat the good of a different, future, person. But what way is this?

It has sometimes been thought that future people's good should count for less than present people's simply because it comes in the future. One should apply a 'social discount rate'. Parfit argues against this view in Appendix F, and I shall not spend time on it. But there are occasions when a future person's good might much more plausibly be thought to count for less. These are when what is in question is whether the future person should ever come into existence at all.

Suppose the government is considering some alterations to the system of family allowances. It is wondering whether to pay more money to larger families. This will encourage parents to have more children. The costs, suppose, will be born by people already alive; perhaps the change is to be financed by reducing pensions. If the change is made, more people will come into existence and live, most of them, worthwhile lives. Should the good they will enjoy be counted as a benefit of the scheme, to be set against the loss to the elderly? It is plausible to think that it should not, or at least that it should count for less than the good of existing people. There seems to be something odd about sacrificing the good of existing people for the sake of bringing new people into existence.

This brings us to consider the size of future generations. Economists often call this the problem of 'optimal population'. It is the subject of Parfit's Part IV. If it was right to draw from Part III the conclusion that a person's future good should be treated a bit like the good of a future person, then this makes an interesting connection between the problem of optimal population and another problem that has troubled economists: what is the value of prolonging a person's life[10]. Prolonging a life is, according to reductionism, a bit like bringing a new person into existence. And we have an analogous question: if a person's life is prolonged, is the whole of the good she goes on to enjoy to be counted as a benefit of prolonging it? The population problem and the problem of valuing life must be treated together.

The population problem has been much discussed in the last fifteen or twenty years. Parfit's earlier work on it has been very influential. Part IV of *Reasons and Persons* draws the discussion together and advances it some more. It is by far the most thorough and authoritative treatment of population ethics that there is. Parfit makes it clear that the difficulty of the subject is easily underestimated, and that no acceptable solution to the problems is yet in sight.

The difficulty is in knowing what the objective ought to be, rather than in finding the best policy for achieving it. In economics until recently only two objectives have been given serious attention: total (or classical) utilitarianism and average utilitarianism. The 'total view' simply aims for the greatest total of good. This has some implausible implications. In the example about family allowances it gives the good of newly created children full weight against the harm done to the elderly. And it leads to what Parfit calls 'the repugnant conclusion'. Suppose the world contains a

[10] I have explored this connection in Broome (1985)

number of people all enjoying very good lives. One can always imagine an alternative world where everyone's life is only just worth living, but where there are so many people that the total of good is larger than it is in the first. According to the total view the second world would be better. To avoid an implication like this many economists have adopted the 'average view' instead. But the average view gets short shrift from Parfit (in Sect. 143). It has very little to recommend it.

Quite recently Jan Narveson (1967) proposed a new approach to these problems. When considering some policy, he suggested, the right thing to do is weigh up how much better or worse off it makes people. If on balance it makes people better off, it is a good policy. Now, if a policy (like increasing family allowances) causes somebody to come into existence, and she has a good life, it is not true that the policy makes her better off. She would have been no worse off without the policy, since she would not have existed at all. So according to this view the good of people who might or might not be brought into existence does not count at all against the good of existing people.

I find this an attractive idea. That a person would be happy if she came into existence does not by itself seem a reason for bringing her into existence. Couples are doing nothing wrong if they decide not to have a child, even though the child would be happy[11]. Parfit, however, deploys a battery of arguments to show that Narveson's idea cannot in the end hold water. I am forced to report that, despite the resistance of my intuition, I find these arguments unanswerable.

An idea like Narveson's motivates Partha Dasgupta's latest work on population (1986). For this reason it is worth looking more closely at one of the difficulties with it. Imagine if you can that the world contains just one person. Three alternatives are available. One is that this person lives alone and her life is good to degree five. The second is that her life is good to degree four, and a second person comes into existence and has a life that is also good to degree four. The third is that the original person's life is good to degree six, and the same second person comes into existence with a life good to degree one. Call these alternatives '(5)', '4,4)' and '(6,1)'. According to the idea I have been mooting (5) is better than (4,4) because it is better for the person who exists anyway. We do not count the good of the second person against this because what is in question is whether she should exist at all. (Even if we gave her good some weight, the result would be the same if the weight was small.) On the other hand (6,1) is better than (5). In comapring (4,4) and (6,1), however, the fact is that both people will exist under either alternative. Choosing (4,4) rather than (6,1) harms the first person and benefits the second. Choosing (6,1) rather than (4,4) does the opposite. There is no reason not to count these harms and benefits to both people equally. Under any reasonable principle (4,4) is better than (6,1). So we have been led to say that (4,4) is better than (6,1), that (6,1) is better than (5) and that (5) is better than (4,4). But logic requires any comparative such as 'better than' or 'hotter than' to be transitive. Our idea, then, has led to a logical contradiction[12].

---

[11] Compare Parfit p. 381

[12] In *Reasons and Persons* Parfit lays little stress on this intransitivity as an objection to principles like Narveson's. He discussed it thoroughly in an earlier paper (Parfit 1976)

What Dasgupta does about this is propose that the choice should be made in two stages. First, he says, one should compare all the alternatives that contain the same number of people, and pick the best. So we compare (4,4) with (6,1) and pick (4,4). Next, having picked the best for each number of people, we compare all of these best together. The best of all these is the one Dasgupta says we ought to choose. So we compare (4,4) with (5) and choose (5). (6,1) is never compared with (5) under this procedure. Dasgupta's procedure, then, in this case succeeds in selecting one of the three alternatives, despite the intransitivity. That, however, is no solution to our problem. Given an intransitivity it would always be possible to set up some procedure for making a choice. Our problem was that the mooted principle led to a contradiction. It still does. The principle still, despite Dasgupta's procedure, says that (6,1) is better than (5). Dasgupta might, perhaps, try to raise the procedure to the level of a new overriding moral principle. He might say that any alternative eliminated by the procedure cannot be better than the alternative chosen. But to justify this he would have to show why his procedure, which happens to make comparisons in a particular order, has this sort of moral force. And in our example he would have to make it convincing that (6,1) is not better than (5). To me at least, it is intuitively clear that it is better.

The outcome of Parfit's complex arguments about population is negative. None of the principles he considers, the objectives one might aim at, proves adequate. He leaves us with the search for 'Theory X' – the right moral theory to guide us in this area – still in progress. But he is not downhearted. Non-religious ethics, he believes, is still in its infancy. We cannot expect it to have solved every problem. And one thing this book shows clearly is Parfit's strong conviction that there must be a right moral theory to be found.

VIII. The Prisoners' Dilemma

Finally I come to Part I. Here Parfit is testing out an argument that has been used against the self-interest theory (the theory that rationality requires one always to act self-interestedly). The argument is that the theory defeats itself by failing to meet its own requirements. Parfit concludes, however, that this argument does not succeed in refuting the theory. He goes on to develop his own argument against the self-interest theory in Part II; I have outlined it in Sect. IV above.

Parfit does agree that in some ways the self-interest theory is self-defeating. It defeats itself, in a way, in prisoners' dilemmas. Prisoners' dilemmas are in practice very common in the modern world, pervaded as it is by externalities and public goods. Here is an example I have gleaned from Alison Booth and David Ulph (1984). Each worker in an industry will find it in her interest not to belong to the union. She will thereby save the subscription, and her belonging would make so little difference to the union's strength that it would not noticeably affect her wage. But if nobody belongs, the wage will be far lower than it would have been if everyone belonged and made the union strong. So everyone will be worse off. Here, then, everyone's acting self-interestedly leaves everyone worse off, in self-interested terms, than they might otherwise have been.

For reasons I shall not go into, Parfit does not think that examples like this are enough to refute the self-interest theory. But they clearly raise a practical problem: if we all behave self-interestedly we shall all be worse off than we might have been. Economists have worried about this problem for a long time. And Parfit devotes some of Part I (Chaps. 2 and 3) to it. Economists have, not surprisingly, generally looked for institutional solutions. These are solutions – externality taxes, laws against pollution and so on – that alter the structure of the situation so that it is no longer a prisoners' dilemma. Parfit, also not surprisingly, looks chiefly for moral solutions. He looks for ways in which changing our motivations could overcome the problem. It would be a mistake to exclude moral solutions from economics. Amartya Sen (1972) has called attention to their practical importance, and also to the mistakes economic theory will make if it ignores them. Booth and Ulph's paper makes it plain how hard it is to explain why people belong to trade unions unless they have some moral motive.

Parfit's contribution is to explain and classify the various possible moral solutions. For instance, we might become reluctant to be free riders. Or we might become altruistic. Or we might become Kantians and act only as we can rationally will that everbody should act. And so on. He particularly spends time on the altruistic solution. The question is: suppose we all become altruists, and act so as to maximize the total of everyone's good rather than just our own, will that overcome the dilemma? The answer is: not necessarily. Parfit devotes a chapter (Chap. 3) to discussing some mistaken arguments that lead to this answer, but these are mostly rather technical mistakes in calculating costs and benefits, of a sort that economists are not likely to fall for. There is, however, a genuine reason why altruism may not solve the dilemma. Suppose that a union of one member would be totally ineffective; some threshold of membership needs to be crossed before the union will have the power to get wages increased. Then if nobody at present belongs, each worker, even is she is altruistic, will see that there is no point in her joining. Nobody will benefit and it will cost her her subscription. Altruism by itself is not enough to bring about the ideal solution where most people belong. This is what Parfit calls a 'coordination problem'. He says a little about solving coordination problems (pp. 72-3, 100-2), but this is a place where there is more work to be done.

IX. Conclusion

Reasons and Persons is one of the most important works of recent moral philosophy. No economist whose work impinges on moral philosophy can afford to ignore it.

References

Booth A, Ulph D (1984) A bargaining model of wages, employment and union membership, Typescript
Broome J (1983) Utilitarianism and separability, Discussion Paper 139/83 of the Department of Economics, University of Bristol
Broome J (1985) The economic value of life, Economica 11:281–94
Dasgupta P (1986) The ethical foundations of population policies, In: Johnson DC, Lee R (eds) Population growth and economic development
Gorman WM (1968) The structure of utility functions, Rev Econom Stud XXXV:367–90
Mirrlees JA (1982) The economic uses of utilitarianism, In: Sen A, Williams B (eds) Utilitarianism and beyond, Cambridge University Press, Cambridge, pp 63–84
Narveson J (1967) Utilitarianism and new generations, Mind LXXVI:62–72
Parfit D (1976) On Doing the Best for Our Children, In: Bayles MD (ed) Ethics and population, Schenkman, Cambridge, pp 100–115
Parfit D (1984) Reason and persons, Oxford University Press, Oxford
Pigou AC (1932) The economics of welfare, 4th edn. Macmillan, New York
Sen A (1972) Behaviour and the concept of preference, Economica XL
Sen A (forthcoming) Consistency, Econometrica