

CONTRACT THEORY AND FUTURE GENERATIONS

Brian Barry

University of British Columbia

Comments (especially on pp 11 ff) wd be welcome. The other is a revision of the earlier one Brian

One way of asking the question what (if any) are our obligations to future generations is to put it in terms of what principles would be agreed on under certain conditions. The objection to this 'ideal contractarian' approach that may be made in general is that we may be more sure what our obligations are than we are sure that we are obliged to do what would be agreed upon in some hypothetical situation. But the force of this objection is substantially weakened in relation to future generations, since it is my impression that most of us have only vague ideas on that question anyway. At the least, the contract framework should help us to clarify our own ideas, even if in the end we discard it.

The obvious point of departure is the powerful and elaborate statement of the 'ideal contractarian' position to be found in Rawls's A Theory of Justice. I shall assume that the outlines of Rawls's theory are familiar. The most important point for the present purpose is that in virtue of the 'veil of ignorance', which conceals from people their personal characteristics, social standing etc., self-interested individuals are forced to agree on principles of general application. They cannot hope to gain special advantages for themselves because they do not know their own distinguishing characteristics.

They do, however, according to Rawls, know that they are all contemporaries. Rawls does not defend this 'present time of entry interpretation' of the 'original position' except to say that the alternatives create difficulties. This is indeed true, as we shall see, but so (as Rawls admits) does the 'present time of entry interpretation'. I shall begin by pointing out the difficulty and arguing that

Rawls's attempted solution to it is not satisfactory.

Rawls's discussion of the whole question of relations between generations is mainly to be found in §44, 'The Problem of Justice between Generations'. The path that Rawls follows here is really quite obscure and in order to be sure just what he is saying we have to scrutinize the text carefully. The initial statement that Rawls makes is as follows:

Thus the persons in the original position are to ask themselves how much they would be willing to save at each stage of advance on the assumption that all other generations are to save at the same rate - that is they are to consider their willingness to save at any given phase of civilization with the understanding that the rates they propose are to regulate the whole span of accumulation. (p. 287.)

This is immediately glossed and we are told it is not a question (as might appear from the above quotation) of picking a single savings rate to which one will adhere 'on the assumption' that other generations will do the same, but rather of choosing 'a just savings principle that assigns an appropriate rate of accumulation to each level of advance'. Apparently there are two criteria: how difficult saving is and how necessary it is. The poorer the society the more important it is to build up wealth but the more hardship saving entails; in a rich society, saving is easier but increasing wealth further is less important. 'Eventually, once just institutions are firmly established, the net accumulation required falls to zero.' Rawls assumes (although it seems to me that nothing he says suggests why the two competing factors have this relationship) that the resultant plot of just saving against GNP per capita will go first up and then down. That is to say, the just savings rate (as a proportion of total national income) will be low at both low and high levels of material advance, and will rise to a maximum at some intermediate level. (See p. 287.)

I shall not ask how plausible is the actual relationship Rawls posits

between the 'just savings rate' and the level of material prosperity. What I want to get at is the idea that the people in the original position are to choose a principle to regulate their savings, whatever kind of society they turn out to belong to, 'on the assumption' that the relevant rate (as determined by the principle they choose) is to be applied by other generations too. But where does this 'assumption' come from?

Rawls says elsewhere that one cannot attribute to the people in the 'original position' false assumptions. However convenient it would be to say that they choose principles on the assumption that their enemy would assign them their place - because the maximin criterion would immediately follow - we are not allowed to do so.

It might be argued in defence of the assumption that others will act on the same 'just savings' principle that, if the members of other generations were also put behind the veil of ignorance, they would choose the same principles. (Indeed, there is no way of differentiating them, since they would have the same information.) But the observation, although true, does not yield the desired conclusion. For the question is precisely what it is rational to choose. The argument that it is rational to make a choice based on the assumption that others will choose the same way because it is rational for the others to make a choice based on the same assumption is as ineffective as any other attempt to raise oneself by tugging on one's own bootstraps.

Precisely this question has often been discussed in relation to the one-shot prisoner's dilemma game. It has been argued (notably by Anatol Rapoport) that it is rational for you to choose the co-operative move because the other player's position is symmetrical with your own, and you should therefore assume that he can see as well as you the joint advantage of both choosing the co-operative rather than the double-cross move. But this reasoning is fallacious because, if

you are self-interested (and if you are not the adjusted pay-off matrix may no longer be of the prisoner's dilemma form anyway) you will always do better to double-cross, whatever the other player does. The only effect of contemplating the symmetry of his position with yours will be to arrive at the melancholy conclusion that he is bound to double-cross you too.

It is really quite surprising how often the logic of the prisoner's dilemma is misrepresented so as to make it appear that the question for a rational self-interested player is whether or not to trust the other player. This would be so only if it were advantageous to oneself to choose the co-operative move when the other player chooses it. But it is not. It is advantageous to oneself if he chooses the co-operative move, of course, but it is advantageous to oneself to double-cross whichever he does. An example of the error may be found in Otomar Bartos's exposition in Simple Models of Group Behavior (New York: Columbia University Press, 1967). He writes: 'We have here a situation in which being rational (in the sense of playing it safe i.e., playing the equilibrium strategy "to confess") leads to a low payoff, and in which ignoring safety and trusting the other player (i.e., by playing "not to confess") could increase one's own payoff' (p. 230). It is simply not true that the rationale for double-crossing is 'playing it safe' (that is to say, a minimax strategy) as against gambling on a higher payoff from an alternative choice. Double-crossing is a dominant strategy and thus indicated by ~~the~~ 'sure thing' principle. There is no way (except as a result of qualms of conscience) in which one could ever regret having double-crossed when one found out what the other player had done.

Rapoport's argument for the rationality of a self-interested person's choosing the co-operative move would be valid only if one player's choosing it were to cause the other player to co-operate. But it is, of course, built into the

specification of the game that the choices are made independently. Similarly, it would solve the inter-generational problem (or at least this bit of it) if the choice of a 'just savings' principle made by one generation actually caused previous generations to act in accordance with its requirements. But the idea of one generation changing the past course of human history by its own choices is, to say the least, mind-boggling. If we stick to the idea that different generations take their decisions independently then, as in the prisoner's dilemma case, the knowledge that others are in a symmetrical position offers cold comfort. Since any given generation are self-interested, and the self-interested line is not to save, they must expect when they emerge from behind the veil of ignorance to discover that their predecessors, acting on the same basis, have not saved either. Rawls's conclusion is, then, as follows: 'Those in the original position know then that they are contemporaries so unless they care at least for their immediate successors there is no reason for them to agree to undertake any saving whatever' (p. 292). He therefore proposes as a solution to the problem to have them care.

The most revealing discussion of this move offered by Rawls comes not in the section on 'The Problem of Justice between Generations' but earlier in §22 'The Circumstances of Justice'. Here he asks 'whether the persons in the original position have obligations and duties to third parties, for example, to their immediate descendants'. He goes on, rather obscurely, as follows: 'To say that they [the parties in the original position] do [have obligations and duties to third parties] would be one way of handling questions of justice between generations. However, the aim of justice as fairness is to derive all duties and obligations from other conditions; so this way out should be avoided.' What this means, I take it, is that we might simply assert peremptorily that the people in the original position have obligations to their descendants, and that would settle the matter. But since the name of the game is derivation from other conditions,

we must try to derive these obligations in the original position from something else.

Now all this is more than a little odd. The object of A Theory of Justice is supposed to be to tell us about the obligations and duties of actual flesh-and-blood people. The point of the construction involving the original position is that (according to Rawls) what would be chosen in the original position constitutes principles of justice. Somewhere along the way, however, Rawls has changed the question and is now asking what obligations ^{and} duties the people in the original position have.

There are two immediate objections to this. The first is that it is surely enough of a job to talk about the obligations and duties of real people, without having to talk about further shadow-obligations of shadow-people - let alone, God help us, derive these shadow-obligations from something else! The second is that it's not at all clear what if anything we would or should or could do with the obligations of people in the original position, even if we had determined what they were. The basic idea of the theory of justice as fairness is, as I said, that justice is comprised by what would be chosen under certain conditions. But if obligations among people in the original position have any relevance they presumably bear on what they ought to choose. I would hate to think where one would go from there.

Fortunately, however, having invited us into this nightmare, Rawls says no more about deriving obligations that are to be imputed to the people in the original position. The tack which Rawls now takes is, he says, to 'make a motivational assumption'. The 'goodwill' of the parties in the original position 'stretches over at least two generations'. We may, though we need not, 'think of the parties as heads of families, and therefore as having a desire to further

the welfare of their nearest descendants'. He concludes as follows: 'What is essential is that each person in the original position should care about the well-being of some of those in the next generation, it being presumed that their concern is for different individuals in each case. Moreover for anyone in the next generation, there is someone who cares for him in the present generation. Thus the interests of all are looked after and, given the "veil of ignorance", the whole strand is tied together.' (All quotes from pp. 128-9.)

This 'motivational assumption' has a desperately ad hoc air about it. If we are to put in the 'motivational assumption' that people care for (some of) their successors it seems difficult to explain why we should not add the 'motivational assumption' that people care for (some of) their contemporaries. Both are equally true of actual people. It may be replied that the second is not necessary for deriving reasonable answers whereas the first is. But this would merely emphasize the ad hoc nature of the move.

More seriously, it is an ad hoc move in that it runs counter to the logic of 'justice as fairness' which is (for better or worse) to derive principles of justice from the constrained pursuit of self-interest. The natural way to deal with the problem of relations between generations within the framework of 'justice as fairness' is to drop the postulate that those in the original position are contemporaries and to say instead that everybody is in the original position, without regard to ~~position~~ in time. Provided nobody knows where he comes in time it would appear that the same logic of self-protection under conditions of anonymity that drives contemporaries to agree on principles of mutual advantage should lead non-contemporaries to agree on principles such that each generation respects the interests of its successors.

This suggestion, however, immediately raises the problem: who is 'everybody'? Presumably different principles may bring into existence different people, different

numbers of people, and make the overall time-span of the human race longer or shorter. The easiest way out is to say that all those who are actually going to exist should be at the convention. This is the line taken by David Richards, who says that 'the class of members of the original position includes, in a hypothetical sense, all persons, who have lived, live now, or will live'. Clearly, this would seem to imply that the 'rational contractors' should do the best for themselves (whether this means maximizing the average, maximizing the minimum, or whatever) and pay no attention to the size or composition of the population that results from their doing so. Richards himself draws this conclusion, and remarks that 'the egoistic desire to exist of the contractors does not influence their consideration of this problem, for ex hypothesi the contractors know they exist in some point of time, and are thus only concerned to ensure that their existence be as satisfying as possible.' (David A. J. Richards, A Theory of Reasons for Action [Oxford: Clarendon Press, 1971_7, pp. 81 and 134, italics in original.]

Surely, however, there is something deeply inconsistent in saying on the one hand that the choice of principles affects the size and composition of the population and on the other hand that the size and composition of the population is given. If we assume (which is not unreasonable) that any principle (or set of principles) will give rise to a different size and composition of population, then any particular size and composition is consistent with only one choice. This surely makes the whole idea of choosing principles collapse.

I can see no way around this problem short of abandoning the postulate that the people in the original position know that they (and only they) actually have existed, exist or will exist. But if we abandon that postulate we have to say that the original position is populated with potential people, some of whom will exist on one set of principles, others of whom will exist on other sets of principles. (There may of course be overlap between the memberships of these alternative

actual populations, and some could be proper subsets of others.)

The implications of this move have been explored by Gregory S. Kavka, in an article entitled 'Rawls on Average and Total Utility' (Philosophical Studies 27 (1975), 237-53). There are two problems here. The first is whether we can derive determinate principles from the notion that they are to be chosen by potential people from behind a veil of ignorance. The second is, if we can, whether we want to attach ethical significance to their choices in the sense that we should regard the principles they would choose as binding on ourselves. Kavka does not address himself directly to the second question. He does say that extending Rawls's analysis to potential people makes it more self-consistent, and I think this is true, but perhaps the conclusion to be drawn from that is that it represents a reductio ad absurdum of the whole idea that obligations of justice may be derived from hypothetical contracts made in peculiar circumstances.

In any case, let us begin by seeing if any sense can be made of the notion that the original position consists of all potential people. The immediate problem that arises is how many potential people there are. The very peculiarity of this question may suggest that we may be heading in an unprofitable direction. As a colleague remarked, it's a fine thing when we give up asking how many angels can stand on the point of a needle in order to ask how many potential people will fit into ~~a meeting hall~~ *the Constitutional Convention*.

If the number is infinite, we had better give up since I very much doubt whether anything can be got out of the idea of an infinite number of people choosing principles which will bring a finite subset of them into existence. But if the number is not infinite it is exceedingly large, and it is difficult to imagine how we would determine its size.

Kavka addresses himself to this question, but does not I think provide an adequate answer.

Let us briefly consider an objection which might arise from those who are willing to countenance possible people in this context. Such persons might object that while there are such things as possible people, there are an infinite number of them and hence that assumption 5 above is mistaken in identifying the number of possible persons as being a finite constant. This objection may be answered in either of two ways. First, we might suggest a physical interpretation for the concept of possible persons such that the constraints of assumption 5 are satisfied. We might, for example, identify possible persons with actual human eggs of which there are certainly a finite number which is larger than the human population which could be brought into existence by even the most dedicated human effort to produce as many people as possible. Second, we may note that treating the constant n as representing the total number of possible persons from which actual populations are drawn, is merely a heuristic device to give backing to the claim that chances of existence are directly proportional to population size. Those who accept possible people and find this device more puzzling than helpful may regard assumption 5 as merely the precise specification of the vague intuitive notion that a larger population means a greater chance that a given possible person will exist. (pp. 245-6.)

The first answer fails to notice that we are not concerned merely with ensuring that n is larger than the total number of people who might ever exist but that it includes all the people who might actually exist. If we follow Kavka in defining 'people' as genetically distinct individuals, we have to recognize that the number is extraordinarily large. Even if the human race had started with just two people - Adam and Eve - their total number of potential descendants would consist of the total number of feasible combinations of one ovum and one sperm. The potential descendants of these potential descendants would consist of all feasible combinations of all the ova of the potential females and all the sperm of the potential males. And so on. Pretty clearly, it would require few generations before the potential people would outnumber the molecules in the universe.

This would not matter too much if Kavka were right in saying that it makes

no difference what number n is so long as it's finite. But it is not in general true that the choice of principles will be unaffected by the size of n . If we want to start off without specific assumptions about the way in which rational people (or more precisely rational potential people) will choose among alternatives, we cannot say that their choice will be unaffected by the size of n , the total number of potential people. For most choice functions, it does make a difference whether one is choosing between two high probabilities or two low probabilities, for example - even if the ratios of the two probabilities are the same in both cases and the values of the alternative outcomes are the same in both cases. In other words, from the fact that someone prefers $p_1 u_1$ to $p_2 u_2$ it does not follow that he must in consistency prefer $p_1 \frac{u_1}{n}$ to $p_2 \frac{u_2}{n}$, where n is any arbitrarily large number. If I am told that a lump sum of a million dollars is to be distributed randomly among a hundred people, including me, and I can choose whether there is to be one prize of a million dollars or ten prizes of a hundred thousand dollars, I will prefer the second. If I am told instead that the million dollars is to be allocated among the entire current population of the world and I am given the same choice I may prefer a minuscule chance of winning the million dollars to an only slightly less minuscule chance of winning a hundred thousand.

If we bear in mind the vanishingly small probability any given potential person has of existing under any possible arrangement, we must surely wonder whether we want to be bound by the choices of these potential people. For them, the choice between different societies is a choice between contingencies all of which are exceedingly remote. Why should the choice between principles to govern interpersonal and intergenerational relationships be determined by the way in which one chooses between alternative contingencies when all of them are vanishingly remote?

Max E(U)

What is true is that n does not matter given certain choice functions, and it so happens that the two upon which Kavka focuses are among these. Maximin - the rule that you maximize the value of the worst possible outcome - has the property that the choice resulting does not depend at all on the probability distribution of outcomes. And the other rule that Kavka considers is the one rule that does take account of probabilities but leaves the ordering of choices unchanged for any n . This is the rule that the expected ^{utility} value is to be maximized, that is to say that, if p_i is the probability that the i 'th outcome will occur and u_i is the utility that will be received from the i 'th outcome and there are n possible outcomes, $\sum_{i=1}^n p_i u_i$ is to be maximized. The general statement of the requirements for maximizing expected utility ~~value~~ can be simplified if it is known that all n outcomes have an equal chance of occurring and that the sum of probabilities is unity. The probability of each outcome is then $\frac{1}{n}$ and the way to maximize expected utility is to maximize $u_1 \cdot \frac{1}{n} + u_2 \cdot \frac{1}{n} + \dots + u_{n-1} \cdot \frac{1}{n} + u_n \cdot \frac{1}{n}$ which may be written $\frac{\sum_{i=1}^n u_i}{n}$. It follows that if n is fixed, $\frac{\sum_{i=1}^n u_i}{n}$ is maximized when $\sum_{i=1}^n u_i$ is maximized. Thus, if each potential person in the original position has an equal chance of being any (existent or non-existent) person, and each is seeking to maximize expected utility, each will agree on the rule that the total utility of all (existent and non-existent) people is to be maximized.

Kavka argues ~~that~~ the value of non-existence should be set as zero. 'Placing the utility of non-existence at the crossover point between net positive utility and net negative utility seems entirely natural since not existing entails experiencing neither happiness nor unhappiness' (p. 241). This is obviously an equivocation: what would be clear would be to say that not existing entails not experiencing either happiness or unhappiness.

If, however, we do give non-existence a value of zero, we can obviously say that the total utility of the non-existent is zero: $\sum_{i=k+1}^n u_i = 0$, where k people exist. Therefore $\sum_{i=1}^n u_i = \sum_{i=1}^k u_i$: the total utility of existent and non-existent people together is equal to the total utility of existent people alone. Total utility for potential people is therefore maximized when total utility for existent people is maximized provided that the maximum total utility is positive. If the maximum total utility is negative, then clearly $\sum_{i=1}^n u_i$ is maximized when $k = 0$, that is to say when no people come into existence. I have not followed Kavka's proof, since he makes several unnecessarily strong assumptions. From (1) maximizing expected utility, (2) non-existence = 0, and (3) total utility attainable by those who exist > 0 , it follows that the rule chosen by the potential people would be to maximize total utility of those who exist.

It may be noted that these are the weakest assumptions that will get the result. If we change (2) to give non-existence a fixed value which is other than zero, it is no longer true that total utility is maximized when the total utility of those who exist is maximized. Instead we have to adjust each existent person's utility by subtracting the utility of non-existence and maximizing the sum of those utilities. The statement of (3) has to be adjusted along the same lines.

This may seem rather trivial, but it is far from it. We now have to realize that the proof has built into it the assumption that a potential person would sooner not be ~~born~~ than be, on balance, slightly unhappy. If we say instead that a potential person would sooner exist than not exist provided he would not suffer more than some certain degree of net unhappiness, we have to add to each existent person's happiness the margin between 0 and the degree of unhappiness that a potential person would just prefer to non-existence. This would mean a larger existent population than would be indicated by the total utility criterion and

a lower total utility of those who exist, because we would go on adding existent people at the margin whose net contribution to total happiness was negative.

The obvious question that arises, however, is whether it is really sensible to treat being a non-existent person as rating a value on a happiness scale that is neutral between happiness or unhappiness - or for that matter assigning it any other value on a scale of happiness. I must say it seems to me clearly a mistake. As I said, it is misleading to suggest that non-existent people experience neither happiness nor unhappiness. Non-existent people don't experience anything. They don't exist. The whole idea of their having claims and interests is mistaken, and there is absolutely no reason why the living should take account of such supposed claims and interests. This is not to say that those living at any given time should not take account of the claims and interests of their successors, but it does mean that the obligation to do so cannot be derived from the idea that we should maximize the expected utilities of the set of all potential people.

The maximin criterion, which Kavka also considers in the context of choice by potential people, produces such bizarre results that it seems to me to cast further doubt on the reasonableness of the criterion itself rather than to offer useful guidance. Kavka adds an extra bizarre note by carrying out the discussion in terms of primary goods. Why maximin should be discussed in these terms when maximizing expected value was discussed in terms of utilities he does not explain. In any case, it is surely clear that defining the maximin principle in terms of primary goods cannot be defended. Once we exist, we want primary goods (let's concede) but it doesn't follow that as potential people we would want to exist in order to have primary goods. As Kavka notes, one has to exist in order to have any of the Rawlsian primary goods. Indeed, he suggests in a footnote (fn. 18, p. 252) that the representative non-existent potential person 'lacks the one fundamental

primary good whose possession is a prerequisite of the possession of all other primary goods - life itself'. Whether or not we treat life as a primary good, however, it may be agreed that 'the representative non-existent person is going to be the worst off with respect to primary goods among all the relevant representative men' (p. 247).

Now the clear implication of this (which Kavka does not recognize) is that on the maximin principle all states of affairs that can actually be achieved are ranked equally. For, as we have seen, the number of potential people is so vastly in excess of the number of people who could exist under any set of arrangements that under any set of arrangements there will be non-existent people. Since all non-existent people are equally badly off in terms of primary goods, the worst-off representative person is equally badly off in all possible states of affairs.

Kavka's own solution is both incoherent and inconsistent with the rationale of maximin as a decision-rule. He continues the sentence quoted above by saying 'and hence the only inequalities justified by the difference principle on this interpretation are those leading to population increases which would increase the expectations with respect to the possession of primary goods of the average non-existent possible person' (p. 247). The point is continued in a footnote (fn. 19, p. 252) which reads: 'These expectations are increased, of course, in accordance with [the assumption of a finite number of potential people], by decreasing the likelihood that a representative person who is non-existent (under status quo social principles) would be non-existent (under the new rules encouraging population growth).'

This is incoherent because any rule produces a set of non-existent people but different rules produce different sets and we cannot therefore speak of trying to improve the prospects of the set of non-existent people. It is inconsistent with the rationale of maximin because the whole idea is to talk about the

levels of possible outcomes and get away from probabilistic expectations. The worst-off potential person (on the primary goods criterion) is one who doesn't exist. We can't get away from that by saying that the worst-off representative potential person can be regarded as existing a little bit because some non-existent people may exist after all.

The conclusion that Kavka reaches - that as many people as possible should be brought into existence - cannot be derived from the argument that this minimizes the number of worst-off people, since the maximin criterion is not concerned with the number of worst-off people. However, we could derive his conclusion from a modification of Rawls's maximin criterion that is a natural extension and has sometimes been suggested. This is the idea that if the worst-off person is equally badly-off in two situations, we go to the next worst-off in each situation, and if one of those is better off than the other we say that situation is preferable. If the next worst-off are equally badly-off we go up one further, and so on until we find a tie-breaking pair. Clearly, it is an implication of this extension of Rawls's maximin principle that if there are in two situations a number of people who are equally worst-off, but there are fewer worst-off people in one than the other, the one where there are fewer is preferable. We can imagine ourselves matching one non-existent person in each situation against a non-existent one in the other, until we finally run out of non-existent people in one. We then match ~~the~~ existent person in that situation with a further non-existent person ~~in the~~ other and declare the first the winner, since something beats nothing in terms of primary goods.

Oddly enough, Kavka does not offer any reflections on his conclusion, which we have seen can be derived validly by an extension of maximin to break ties. Yet surely it is simply insane to suppose that we ought to maximize the total number of people ever born so as to give as many potential people as possible the

chance of getting some primary goods (even if little more than life itself). The absurdity of trying to serve the interests of potential people is surely brought out starkly by the conclusion. It is perhaps if anything brought out even more starkly by the implication of the unmodified Rawls maximin criterion, which is that all states of affairs are equally just, however the people alive in them fare, because there are some non-existent potential people in all of them.

Does the craziness of the conclusion follow from the use of primary goods rather than utilities? Certainly, the conclusion is altered if we switch to utilities, but in a way that I am inclined to think casts even further doubt on the maximin criterion. Before saying what that conclusion is, I should note that Kavka, in the paragraph following the one I have been quoting, restates his own conclusion in a way that involves utilities, though he does not draw attention to the shift. He now says that what would be chosen would be 'some different set of principles [from those put forward by Rawls] designed to strongly encourage and facilitate population growth up to the point where limits on natural resources make life not worth living for the average citizen' (p. 247). However, as a statement of the implications of maximin for utilities this again fails because of an illegitimate use of averaging. Whether or not the average person who is alive is better off alive than non-existent is neither here nor there for a potential person who is following a maximin criterion. The question he has to ask himself is what the worst possible outcome would be under various alternatives. (This may be made clearer if it is recalled that the person in Rawls's 'original position' is supposed not to worry about the average obtainable under alternative arrangements but to concentrate on the minimum he might obtain.)

Let me offer what seems a safe statement. There is no imaginable rule for the conduct of human affairs that would not result in there being at least one person in the history of the human race who would regret having been born. If

we are prepared to attach a value to non-existence, then I take it this must mean that - whether the value of non-existence is positive, negative or zero - it is possible to be worse-off by being alive than by being non-existent. As I have already remarked, I am far from happy with the whole notion of attaching a value to non-existence, but unless we are prepared to countenance it at least in order to see where it gets us I do not see how the business of ascribing choices to potential people who may exist or not can be got off the ground at all.

If I am correct, then, in saying that some human life has a value lower than non-existence under any set of arrangements, it immediately follows what decision potential people following a maximin rule will take. They will opt not to bring the human race into existence. For it is clear that the worst possible outcome that one might obtain from existing is worse than the guaranteed outcome from not existing. There is, of course, no reason why we should not embrace this as the correct answer if we choose. But if we find the answer unreasonable, we have a further case for thinking that the maximin criterion is no better as a guide to making large decisions than it is as a guide to making smaller ones.

October 1975