Welfarism and Utilitarianism: A Rehabilitation*

YEW-KWANG NG
Monash University

Utilitarianism seems to be going out of fashion, amidst increasing concerns for issues of freedom, equality, and justice. At least, anti-utilitarian and non-utilitarian moral philosophers have been very active. This paper is a very modest attempt to defend utilitarianism in particular and welfarism (i.e., general utilitarianism or utilitarianism without the sum-ranking aspect) in general. Section I provides an axiomatic defence of welfarism and utilitarianism. Section II discusses the divergences between individual preferences and individual welfares and argues in favour of welfare utilitarianism. Section III criticizes some non-utilitarian principles, including knowledge as intrinsically good, rights-based ethics, and Rawls's second principle. Section IV argues that most objections to welfarism are probably based on the confusion of non-ultimate considerations with basic values. This is discussed with reference to some recent philosophical writings which abound with such confusion. Section V argues that the acceptance of utilitarianism may be facilitated by the distinction between ideal morality and self-interest which also resolves the dilemma of average versus total utility maximization in optimal population theory.

I. WELFARISM AND UTILITARIANISM

A. *An Argument for General Utilitarianism*

Roughly speaking, general utilitarianism (or welfarism) is the belief that what makes anything (an act, a change, a policy measure, a rule) morally right or wrong, good or bad, *ultimately* depends only on its effects on individual utilities.[1] Utilitarianism in its specific or classical form further says that what is important is just the sum of individual utilities. In terms of an objective function or social welfare function, we have $w = w(u^1, u^2, \ldots, u^n)$ for welfarism and

* I am grateful to Peter Singer for commenting on the first draft.

[1] In this section, we ignore the question what individual utilities stand for. In particular, whether they refer to individual preferences or individual happiness (discussed in section II); whether non-human sentients like animals are included in the relevant set of individuals. See Y. K. Ng, 'Some Broader Issues in Social Welfare', in P. K. Pattanaik and M. Salles eds., *Social Choice and Welfare*, Amsterdam, 1983.

© Oxford University Press 1990

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

$$w = u^1 + u^2 + \ldots + u^n \equiv \sum_{i=1}^{n} u^i$$

for utilitarianism, where w = social welfare, u^i = utility of individual i, and n = the number of individuals.[2] Obviously, utilitarianism presumes welfarism. In other words, a utilitarian must logically be a welfarist but not necessarily vice versa. In its unqualified form stated above, welfarism (including of course utilitarianism) subsumes consequentialism—the belief that the moral goodness of an action depends only on its consequences. However, it is possible to divorce welfarism from consequentialism, as discussed below.

A main objection to consequentialism is that the motive behind an action should not be completely ignored in assessing the moral goodness of an action. Two different reasons for this belief may be distinguished. First, good motives and their encouragement may contribute to good consequences in the long run. But this is not inconsistent with consequentialism provided a long-run, complete view is taken.[3] Secondly, some people may regard good motives as good in themselves irrespective of their consequences. Person A attempted out of altruism to save someone dying at a great cost to himself. However, due to ignorance, he used a wrong method which hastened the death. Person B maliciously tried to kill someone just for fun. However, due to some unlikely occurrences, some benefits were effected. Most people feel that A is good and B is bad rather than the other way round. This is partly because we know that A/B is likely to do many other good/bad deeds most of which would lead to good/bad consequences. Even abstracting from this factor, most people still believe that motives cannot be completely ignored in assessing moral goodness either of a person or of an action. This objection to consequentialism may be resolved by the following distinction.

Let us apply moral goodness to motives and to consequences separately. We can then say that the motive behind action X is good though its consequences are bad or vice versa. We may then confine welfarism to moral assessment with respect to consequences. Thus revised, welfarism is the belief that what makes anything morally good or bad in terms of its consequences ultimately depends only on its effects on individual utilities. In other words, welfarism rules out non-utility consequences as of ultimate relevance for moral goodness with respect to consequences.

[2] I use 'welfarism' as a synonym for 'general utilitarianism', see A. K. Sen, 'Personal Utilities and Public Judgements', *Economic Journal*, lxxxix (1979), 537–58. However, some readers may prefer to use welfarism as $w = w(w^1, w^2, \ldots, w^n)$ where w^1 = welfare or happiness of individual 1. To be precise, $w = w(u^1, u^2, \ldots, u^n)$ is preference welfarism, $w = w(w^1, w^2, \ldots, w^n)$ is welfare welfarism.

[3] C. Radford, 'Utilitarianism and the Noble Art', *Philosophy*, lxiii (1988), 63–81.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

Consider the following value axiom.

A1. *In a world with no conscious beings, nothing is of any moral consequence*

Here, a 'world' is defined as a four dimensional totality, encompassing all space and all times and of course all beings and events therein to eternity. Thus a world with no conscious beings means no conscious beings even in the future. Possible goodness such as in the conservation of resources or natural beauties for possible future beings to appreciate does not therefore arise. Since there is nothing conscious in such a world to perceive anything, it seems extremely reasonable, if not compelling, to say that nothing is of any moral consequence.

From A1, it does not take a huge step to accept A2.

A2. *In a world with no sentients, nothing is of any moral consequence*

Here, a sentient is defined as a being that is capable of being happy or miserable. Thus, a sentient must be a conscious being but a conscious being need not be a sentient. For example, one can logically imagine beings that can perceive lights or even see and feel (the existence of) things without being able to feel subjective pleasure and pain. In fact, many lower forms of animals in our world are likely to actually fit this description. It is likely that the ability to perceive evolved earlier than the ability to feel rewards and punishment, because the latter ability serves no purpose without the former.[4] However, while these animals eventually evolve into sentients in our world, they shall never do so in a world satisfying the precondition of A2.

Suppose that our world W_0 is transformed into W_1 by changing all sentients (including ourselves) into non-sentient conscious beings (to eternity). According to A2, nothing is of any moral consequence in W_1. However, we (i.e., human beings after being made non-sentient but still conscious) may still regard some acts (such as lying) as immoral though without the feeling of moral compulsion (otherwise we would still be sentients). But this must surely be due to our memory of affairs in our present world. Let us transform our present world W_0 into W_1 by not only changing all sentients into non-sentient conscious beings but also eliminating their memories of affairs in W_0. Would lying still be regarded as immoral in W_1? Many people (including say Arthur) may

[4] However, perception and affective feelings evolved simultaneously. As I hope to argue elsewhere, perception without affective feelings also serves no evolutionary purposes. Here, perception is defined as involving subjective consciousness, not just in the wider sense 'reception and interpretation of signals from the environment', regarded as evidenced in the bacterium E. coli, see M. Delbruik, *Mind from Matter? An Essay on Evolutionary Epistemology*, Oxford, 1986.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

reply in the affirmative and believe themselves to have such a moral view in W_1. To see that such a view is untenable, consider a different world W_2 which is similar to ours in most aspects (it even has a counterpart of each and every individual in our world) except that lying in W_2 does not cause any real hardship, mistrust, or any other undesirable effects. Rather they lie to add colours to their life. People that seldom lie are regarded as rather dull. Lying in W_2 is like telling a joke in W_0. Everyone in W_2 regards lying as morally good. Now, let us transform W_2 into W_3 by making all sentients into non-sentient conscious beings and eliminate all their memories of affairs in W_2. Moreover, make some minor changes if necessary such as to make W_3 exactly the same as W_1 in every aspect. If we ask the counterpart of Arthur in W_2 (call him $Arthur_2$) what he thinks or would think of the moral standing of lying in W_3, he will probably say that lying is still morally good even in W_3. But by construction W_3 is identical to W_1. How could the same thing (lying) in the world (W_1 or W_3) be both morally bad and good at the same time? Such absurdity can be avoided by accepting A2.

The acceptance of A2 seems to compel the acceptance of A3.

A3. *In a world with sentients, what is of moral consequence is ultimately due to the effects on the feelings of happiness or unhappiness of the sentients in that world*

If what is of moral consequence is not ultimately due to the effects on the feelings of happiness or unhappiness of sentients (or sentient feelings for short), then something may be of moral consequence even without in any way, directly or indirectly, affecting sentient feelings. If this is true, something should also be of possible moral consequence in a world with non-sentient conscious beings since there is no morally relevant difference between the two. But this violates A2. Hence we must also accept A3. Using utility to represent sentient feelings, A3 is welfarism. Ultimately, what is morally good or bad depends only on the effects on individual utilities,

$$w = w(u^1, u^2, \ldots, u^n).$$

B. *From General Utilitarianism to Specific Utilitarianism*

General utilitarianism in the form $w = w(u^1,u^2, \ldots, u^n)$ allows for many specific forms of the function $w(u^1 \ldots . u^n)$, and the Classical form of unweighted summation is only a specific possibility. However, classical utilitarianism has been derived under very reasonable sets of

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

axioms. First Harsanyi[5] showed that under the axioms that both individual and social preferences satisfy the Marschak postulates for expected utility maximization and that society is indifferent if all individuals are indifferent, social welfare is a weighted sum of individual utilities. Recognizing finite sensibility, Ng[6] showed that social welfare is an unweighted sum of individual utilities mainly under the axiom of Weak Majority Preference: the society prefers x to y if no individual prefers y to x and at least half of the individuals prefer x to y.

Harsanyi's and Ng's derivation of utilitarianism differ from some other axiomatic derivations of utilitarianism.[7] The latter are based on informational restriction which *rules out* non-utilitarian social welfare functions. Utilitarianism then follows from the assumption of the existence of some social welfare function. This is similar to Arrow's impossibility theorem in the following sense. Assuming A, B, C, the social choice rule must be dictatorial. That does not mean we want dictatorship which may be worse than non-A, non-B, and/or non-C. In contrast to these 'negative' approaches, Harsanyi's and Ng's results are positive. Their axioms positively imply utilitarianism.[8]

Harsanyi obtained only *weighted* summation. However, unweighted summation can be obtained by adding an additional axiom of anonymity. Each individual's utility, provided measured by some common or interpersonally comparable unit (the finite sensibility approach is one method for obtaining this), should be treated anonymously and similarly.

II. WELFARE UTILITARIANISM VERSUS PREFERENCE UTILITARIANISM

Both Harsanyi's and Ng's results refer to individual utilities (which represent preferences, in contrast to welfare which I define as happiness). When individual preferences and individual welfares do not differ, it does not matter whether one has preference utilitarianism or welfare utilitarianism. In general, preference and welfare may differ for a number of reasons.

First, preference may differ from welfare due to ignorance and

[5] J. C. Harsanyi, 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility', *Journal of Political Economy*, lxviii (1955), 309–21.

[6] Y. K. Ng, 'Bentham or Bergson? Finite Sensibility, Utility Functions, and Social Welfare Functions', *Review of Economic Studies*, xlii (1975), 545–70.

[7] C. d'Aspremont and L. Gevers, 'Equity and the Informational Basis of Collective Choice', *Review of Economic Studies*, xliv (1977), 199–209.

[8] Fleming's result is also positive but he established only separability, i.e., $w = f^2(u^2) + \ldots, + f^n(u^n)$, but using very reasonable axioms. M. Fleming, 'A Cardinal Concept of Welfare', *Quarterly Journal of Economics*, lxvi (1952), 366–84.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

imperfect foresight. For divergence due to this, most people (including preference utilitarians) will agree that it should be welfares that enter the social welfare function. One would not want to maximize something due to mistakes. It may be sensible to ignore such divergence and to maximize social welfare as a function of individual preferences even if the latter are based on some mistakes, due to the practical reason that to do otherwise would lead to greater harm, e.g., fostering illiberal policies. However, this only means that we are really maximizing long-term social welfare (still a function of individual welfares) by using the method of maximizing the short-term function of individual preferences.

Secondly, preference may differ from welfare due to a concern for the welfares of others. Thus it is possible for a person to prefer x to y and yet be himself less happy in x than in y because he believes, e.g., that other people are happier in x than in y. It is true that the belief that other people are happy may make him happy. But this may not be strong enough to outweigh the loss he has to suffer from changing from y to x. For example a person may vote for party x, knowing that he himself will be better off with party y in government. The reason he votes for x is that he believes that the majority of the people will be much better off with x. This itself may make him feel better and is a form of external effect. However, this external benefit may not be important enough to overbalance, in terms of his subjective happiness, his personal loss, say in income, under x. He may yet vote for x due to his moral concern for the majority. To give an even more dramatic example, consider an individual who expects to lead a very happy life. When his country is being invaded, he may volunteer for a mission which will bring him the certainty of death. The prospect of being a citizen of a conquered nation especially with the guilty conscience of failing to volunteer for the mission may not be too bright. But overall he may still expect to be fairly happy leading such a life. Yet he chooses death for the sake of his fellow countrymen. He is not maximizing his own welfare.

When preference differs from welfare due to the concern for the welfare of others, it seems clear that it is welfare, not preference, that should be taken account of in our social welfare function. The reason is simple. I (indexed k) have the preference function $u^k = w^k + \alpha\sum_{i\neq k} w^i$ where α is positive number smaller than one and $\sum_{i\neq k} w^i$ is the sum of all individual welfares other than my own welfare w^k. Not being perfectly self-concerning, my α is larger than zero. Not being perfectly moral, it is less than one. (But I refuse to disclose its exact value. My disclosure might astound many moral philosophers and economists; the former thinking it unethically small and the latter finding it incredibly large.) If the welfare of others (w^i, $i\neq k$) are already taken into account on a

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

par with mine in an anonymous social welfare function $w = w(w^1, \ldots, w^n)$, I would not want my partial concern for others $\alpha_{i \neq k}^{\Sigma w^i}$ to be taken into account on top of $w^1, \ldots, w^n$, since all w^i, $i \neq k$ are already fully taken into account.

I lump all divergences of preference from welfare other than the above two sources (imperfect knowledge and concern for the welfare of others) as irrational preferences. (Some people may prefer to call this welfare-irrational.) Some irrational preferences (e.g., inertia, lack of will-power) are recognized or admitted by the individuals concerned to be irrational or unwarranted. For example, once seated comfortably in front of the TV set, many people prefer to continue watching TV even though they admit that they may be better-off getting up to do more important things and that they would be better-off by more than the small cost of having to get up from a comfortable position. Another source of 'recognized' irrationality involves choices made or preferences expressed during time of somewhat abnormal state of mind such as emotional upset, stress, drunkenness, ecstasy, etc. The person concerned may realize his irrationality when in a more normal state of mind. It is clear that 'recognized' irrational preferences are mistakes not very dissimilar from preferences based on imperfect knowledge. Hence, it is again individual welfares that should affect social welfare.

For 'unrecognized' irrational preferences, it is more controversial. While they may be defined as irrational according to my definition, the individuals concerned may not agree that they are *really* irrational. For such 'unrecognized' irrational preferences, I will certainly agree on the practical ground of liberalism that it is probably best to respect all individuals and accept their preferences as argument in the social welfare function even though they differ from individual welfares. However, on a more ultimate or fundamental level, I think welfares are more appropriate arguments in the social welfare function. Nevertheless, if some individual insists on having his preference instead of his welfare to be taken account of, I will not object after some persuasion for his own benefit.

Why do I regard welfare as a rational and the only rational ultimate objective? I know that my welfare or happiness is of value to me for its own sake because I enjoy my pleasant feelings and suffer from pain. I also observe and understand from other people's admission that virtually all other people want their own happiness for its own sake. By imputation, I think it is rational (at least highly moral) also to want happiness for others. Thus, it seems clear that happiness or welfare is a rational ultimate objective. Why the only one? Because I have never found any other acceptable *ultimate* objective. The next section discusses why some common candidates for ultimate goodness do not stand up to scrutiny.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

III. A CRITIQUE OF SOME NON-UTILITARIAN PRINCIPLES

Many non-utilitarian candidates for ultimate goodness have been proposed. It is however unnecessary to be exhaustive as similar arguments against some are applicable against others.

A. *Knowledge as an Ultimate Goodness*

I accept that knowledge is good but this is so only because in most cases knowledge tends to increase the sum total of happiness, at least in the long run. But is knowledge a good thing in itself independent of its contribution to welfare?

Apart from its instrumental utilitarian values, most people regard knowledge as good because, I believe, we are genetically programmed to be curious, making us want to acquire knowledge. The survival of *homo sapiens* depends largely on its superior knowledge in comparison to other species. Curiosity is thus a trait beneficial to survival. We are not only curious to acquire knowledge ex-ante, we also derive much satisfaction in acquiring knowledge ex-post. That a survival-enhancing trait should be rewarded is of course a common genetic principle. Thus, apart from its more instrumental values, knowledge also increases our welfare by satisfying our curiosity instinct ex-ante and ex-post. If this is what is meant by 'knowledge is good in itself', I fully agree but then it is not inconsistent with using welfare as the only criterion of ultimate goodness. On the other hand, if 'knowledge is good in itself' is meant to be independent of its utilitarian instrumental values and of its direct contribution to welfare due to our satisfaction of curiosity, and of any other possible welfare–related reasons, then I find it unacceptable.

Consider the following example. A world of two peoples living in two planets light years apart. People A is very prosperous and happy, and people B poor and happy. Suppose it is absolutely certain that, due to the great distance, neither can help the other nor learn anything useful from the other. The relevant increase in knowledge is learning of the existence and conditions of the other people. People A, if they know of people B, would become miserable by being unable to help them. People B would also become miserable to learn of their poverty relative to A. This increase in knowledge is clearly bad.

It may be argued that the above increase in knowledge is good in itself but the goodness is overbalanced by the undersirable side effect of a decrease in welfare for both peoples. This is a position of a pluralist that both knowledge and welfare (or some other candidates) are good

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

in themselves. I find pluralism unsustainable. (This will be argued in a separate paper.) Here, from the viewpoint of welfarism, it may be said that, even if people A and people B are not made miserable but just a little worse-off (in welfare), I will regard the increase in knowledge as undesirable. Assuming that peoples A and B are similar to us in respect of curiosity, they will be made better off by the increase in knowledge as such. But if the above-mentioned undesirable effects more (no matter by how much) than offset the positive effect to produce an overall negative welfare effect (taking all effects into account), I regard the relevant increase in knowledge as undesirable. If the overall effect is zero, I am indifferent.

That the relevant increase in knowledge is undesirable does not mean that it is desirable to obstruct it from taking place since doing so generally has far-reaching undesirable effects. Due to the general positive effects of knowledge on welfare and to the negative effects of the restriction of knowledge, it may be desirable to adopt a general principle of freedom of acquiring and distributing knowledge despite that, in some specific instances, the increase in knowledge may be undesirable.

B. *The Unacceptability of Rights-based Ethics*

Many people believe in certain 'rights' (e.g., human rights, political rights, rights to life, free speech, free contract, etc.) which are independent of and antecedent to the consequences (utilitarian or otherwise) of the exercise of these rights. This can be shown to be an unacceptable position to take.

As a matter of fact, I am in favour of most of the commonly upheld 'rights' *because* the adherence to them generally promotes the general welfare.[9] But viewing these (or any other) rights as independent of and antecedent to their welfare implications is absurd. It may even be politically and practically desirable to treat these rights *as if* they are sacred and hence independent of, antecedent to, and even more important than the general welfare. However, why this is so is again because it is beneficial to the general welfare to do so. Thus, on the fundamental philosophical level of ultimate justification, no rights should be independent of welfare. Most rights may be good practical

[9] See D. Lyons, 'Human Rights and the General Welfare', *Philosophy and Public Affairs*, vi (1976), 113–29; R. Hardin, 'The Utilitarian Logic of Liberalism', *Ethics*, xcvii (1986), 47–74; and A. Kuflik, 'The Utilitarian Logic of Inalienable Rights', *Ethics*, xcvii (1986), 75–87 on the consistency of utilitarianism with certain inalienable rights. However, the preference for certain counter-efficiency rights may be based on the ignorance of economics; see Y. K. Ng, 'Economic Efficiency versus Egalitarian Rights', *Kyklos*, xli (1988), 215–37.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

morality and politics, but *never* good moral philosophy at the ultimate level.

One way to see the unacceptability of welfare-independent rights is to ask the question '*why* Right X?' to a very ultimate level. If the answer is 'Right X because Y', then one should ask 'Why Y?'. For example, if the answer to 'why free speech?' is that people enjoy free speech, it is already not welfare-independent. If the answer is 'Free speech deters dictatorship', then we should ask, 'Why is it desirable to deter dictatorship?' If one presses hard enough with such questions, most people will eventually come up with a welfare-related answer.

A second way is to consider cases where two or more rights or principles are in conflict with each other. For example, a broadcaster in Melbourne was sacked by a radio station for making racist comments. Acknowledging the principle of free speech, a spokesman for the station said that certain limits are required. Presumably, free speech conflicts with non-racism. There are different degrees of racist remarks. If there is only a slightest hint of racism, one may be willing to uphold free speech. If it is outright racist outcry, many people may be in favour of the sanction. How are we to decide what degree of racism is not tolerable by free speech? The ultimate damage to welfare can provide a consistent criterion for the appropriate reconciliation of conflicting rights and moral principles.

Thirdly, questioning the scope where certain rights are applicable may also be insightful. For example, consider the right of animals to freedom from cruelty, e.g., being kicked purposefully. Why is this right applicable only to animals, why not also to vegetables, rocks, chairs, etc.? Obviously, this is because animals feel pain (negative welfare) if kicked while vegetables, etc. do not. The right is not welfare-independent.

Fourthly, considering circumstances or examples where the exercise of certain rights leads to disastrous outcomes in terms of welfare will also likely persuade us of the unacceptability of welfare-independent rights. Before providing a specific example, it may be noted that ultimate moral principles are meant to apply under all conceivable (even if extremely unlikely or even impossible, but not illogical) circumstances. Thus counter-examples cannot be dismissed on the ground of their hypothetical nature.

Suppose the Devil offers an individual the option of extending his life by ten years (or twenty, or a certain specified improvement in health) in exchange for becoming sterile (but sexually still active). The right of the individual to freedom to enter into mutually agreeable contract requires that he be allowed to accept the Devil's offer if agreeable. However, suppose the Devil strikes such a contract with each and every individual on earth, the human race will be extinct in about a

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

hundred years. Assuming that on average human beings enjoy positive welfare which may also be expected to increase in the future with the advancement of science and with economic growth. The extinction thus denies substantial positive welfare to future generations for a long time at a relatively much smaller benefit of a slight prolongation of life for all individuals now. This is a disastrous outcome. Under this and other circumstances where individual freedom to contract will result, through either ignorance, lack of co-ordination, or any other factor, in disastrous outcomes, I do not think that the right to free contract should be upheld, unless not upholding this right would be more disastrous.[10]

One cannot take the position that one will stick to a certain welfare-independent Right X *except* under circumstances where it leads to welfare-disastrous outcomes. This position is illogical because if Right X is truly welfare-independent, it should not depend on its outcomes, however disastrous. Thus, if one cannot accept certain rights under any conceivable circumstances, one has to reject those rights as one's *basic* moral principles.

It must however be recognized that once certain rights or moral principles are accepted, even initially for their promotion of welfare, they tend to be, in time, valued for their own sake by most individuals. Their violation or just suspension may thus be disturbing to us. The loss in welfare due to this has to be taken into account. Also, the suspension of one right or principle to avoid a welfare-disastrous outcome may also tend to make other (still welfare-enhancing) rights and principles less sacred. This may have the disadvantage of reducing their observance and thus reduce welfare. This and all other relevant effects have also to be taken into consideration in the decision to suspend or cancel a given right or principle.

Our sentiment of justice may also be partially explained by its survival value.[11] However, what contributes to survival need not necessarily contribute to welfare; our preferences and actions influenced by such genetic factors may thus lead to welfare-irrational choices.[12] To maximize welfare, we may have to learn to make some adjustments against our instinctive sentiments.

[10] For some more realistic circumstances where the exercise of rights leads to disastrous outcomes, see A. K. Sen, 'The Moral Standing of the Market', in E. F. Paul *et al* eds., *Ethics and Economics*, Oxford, 1985, p. 6.

[11] A. Gibbard, 'Inchoately Utilitarian Common Sense', in H. B. Miller and W. H. Williams eds., *The Limits of Utilitarianism*, Minnesota, 1982, p. 83.

[12] Y. K. Ng, 'Individual Irrationality and Social Welfare', *Social Choice and Welfare*, vi (1989), 87–102.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

C. *A Critique of Rawls*

If all welfare-independent rights are unacceptable as fundamental moral principles, Rawls's second principle of justice (maximin) is absurd.[13] (It is, however, surpassed in absurdity by maximax discussed in Section IV below).

The first principle requires that each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all. I am prepared to accept this first principle on the following understanding:

(1) that it is adopted *because* it promotes the general welfare;
(2) in circumstances where it is disastrous to the general welfare, it may have to be suspended;
(3) in deciding what is the 'most extensive total system' and what is 'compatible ... for all', the ultimate criterion is the general welfare.

A sex maniac may be in favour of freedom to rape and claims that this is compatible with everyone's freedom to rape.[14] It may also happen that the sex maniac is a person of the lowest welfare level such that freedom to rape for all will maximize the welfare of the worst off, hence being consistent with the spirit of Rawls's second principle to be discussed below. However, if freedom to rape results in the reduction of the welfare of those raped and scared of being raped by more than (in aggregate) the welfare gain of the rapists (even though the former still have higher welfare levels than the latter group even with freedom to rape), then freedom to rape should be regarded as not compatible with the freedom of not being raped. The 'most extensive total system of basic liberties' should then not include the freedom to rape. However, thus interpreted, the first principle is really a device to promote the general welfare. It is not ultimate.

Despite its obvious absurdity, Rawls's second principle is very popular. For example, Temkin believes that, in 'one form or another, many philosophers have come to advocate a maximin principle of justice, and one can see why. There is strong appeal to the view that just as it would be right for a mother to devote most of her effort and resources to her neediest child, so it would be right for society to devote most of *its* effort and resources to its neediest members.'[15] In my

[13] J. Rawls, *A Theory of Justice*, Oxford, 1972, pp. 150–61.

[14] It is true that Rawls would argue that freedom to rape is not a basic liberty while the right to non-violation of the body is. However, how do we determine what are basic liberties? Either it is based on the utilitarian principle or it is open to the objection of the last subsection on the unacceptability of rights-based ethics.

[15] L. S. Temkin, 'Inequality', *Philosophy and Public Affairs*, xv (1986), p. 109.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

view, the ethical appeal of this argument as well as that of the maximin principle of justice itself is not difficult to question.

I agree that, in most cases, a mother should devote more, and in many cases, most of her effort and resources to her neediest child, but only because this maximizes the welfare of the whole family. The most disadvantaged child is usually the neediest one because he/she will suffer most in the absence of extra help. Thus, the extra care for the most disadvantaged need not be inconsistent with overall welfare maximization. However, the maximin principle requires the mother to go much further.

For simplicity, suppose that the mother is faced with only two alternatives. One is to go away with the most disadvantaged child to live in a mountain resort for certain marginal benefit to the health of the child. The other is to stay at home looking after all the five children but still with possibly more care for the most disadvantaged one. Suppose the two welfare profiles for the children are

$WP_A = (10, 10, 10, 10, 9)$

$WP_H = (900, 900, 900, 900, 8)$

and that the mother is indifferent herself and no one else is affected by the choice. The maximin principle requires choosing WP_A. This, in comparison to the alternative of staying at home, increases the welfare of the worst-off child from 8 to 9 at the costs of a huge reduction in welfare (from 900 to 10) for every child. No sane mother in the world would make such an absurd choice.

Note that the welfare profiles WP_A and WP_H above are ultimate outcomes, as must be the case for all discussion of the ultimate ethical principles. Thus, the more equal welfare profile of WP_A would not promote a further gain in welfare through, say, a more harmonious family relationship. Such effects, if any, should already have been incorporated into WP_A and WP_H.

Consider the much-cherished principle, 'From each according to his ability; to each according to his needs' (which I personally approve, assuming no disincentive effect). Why doesn't it read, 'An equal amount of work from each; an equal amount of income to each'? If a weak man is tired by four hours of work, it is better for a stronger man to work longer to relieve him. Similarly, if the worst-off child will not gain much more happiness by extra effort and resources, it is better that these resources be spent on other children.

In the original position, behind the veil of ignorance as to which child I would be, I would have not the slightest hesitation in wishing my mother to maximize the welfare of all the children, i.e., the unweighted sum of all children's welfare (the welfare of the mother and that of any other person are being held constant in this comparison). This maximizes my as well as other children's expected welfare.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

It is sometimes argued that a risk-averse person may not want to maximize expected welfare. I think it is quite rational to be risk-averse with respect to income or any other objective rewards since one may have, with good reasons, diminishing marginal utility/welfare of income. But since utility/welfare is the ultimate objective one is presumably maximizing, it is not rational not to maximize expected utility/welfare, if the relevant utility/welfare profiles already included all relevant effects, including such things as anxiety, excitement, etc. which explain most paradoxes of choices involving risk such as the Allais.[16] Secondly, if one chooses to be risk-averse with respect to welfare, it is still impossible to justify reasonably an *absolute* degree of risk-averseness as implied by the maximin principle.

After lifting the veil of ignorance, I would still think that it is right for my mother or anyone's mother to maximize the welfare of all children together whether I were the worst-off child or not. Again, my bias in favour of my own welfare may mean that I would hope that my mother would spend more effort and resources on *me* somewhat beyond the level justified by unweighted sum maximization. (But I think it is unjust for my mother to follow the partial wish of any child.) However, even then, I would definitely *not want* my mother to maximin in my favour, implying a zero trade-off on the welfares of my brothers and sisters as long as their welfares remain higher than mine. Thus, given the welfare profiles WP_A and WP_H above, even if I were the worst–off child, I would want my mother to choose WP_H. It may be thought that since my welfare is higher in WP_A, I could not have wanted my mother to choose WP_H. This ignores the differences between welfare and preference due to a concern for the welfare of others as discussed in Section II above.

From the above, it may be concluded that the maximin principle of justice is not only utterly unacceptable but an ethical principle similar to the one in favour of its adoption seems to require *the worst-off group* itself not to accept it. Why then is the principle so popular? One possible explanation is that it appeals to the guilt feeling of the better-off. They have admirable sympathy for the worst-off but yet are not prepared and/or find it ineffective to alleviate this by substantial personal contribution to the worst-off (by charity or the like). Paying lip-service by advocating the maximin principle of justice is a much more cost-effective way of alleviating their sense of guilt. Of course, this explanation need not apply to all advocates of maximin.

[16] M. Allais and O. Hagèn, *The Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, 1979. See also J. Harsanyi, *Essays on Ethics, Social Behaviour and Scientific Explanation*, Dordrecht, 1976, and Y. K. Ng, 'Expected Subjective Utility: Is the Neumann-Morganstern Utility Index the same as the Neoclassical's?', *Social Choice and Welfare*, i (1984), 177–86.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

IV. THE FALLACY OF CONFUSING NON-ULTIMATE CONSIDERATIONS WITH BASIC VALUES

Most objections to welfarism in general and utilitarianism in particular are based, I believe, on the confusion of non-ultimate considerations with ultimate (or basic) values. This confusion is also responsible for many other controversies in moral philosophy. This belief is reinforced by a review of recent philosophical literature. Examples can easily be multiplied but a couple are sufficient to illustrate the point.[17]

Temkin[18] considers Diagram 1 where the width of the blocks represents the number of people living, and the height, their welfare or quality of life. In comparison to A, B involves a population twice as big, and similarly for D in comparison to C. Since proportional changes in population size alone do not affect inequality, the degree of inequality is identical in A and B, and also identical in C and D.

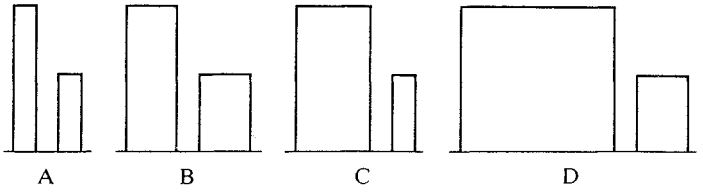

Diagram 1.

Most people would agree that C is better than B in equality, i.e., that B would be improved if some of the worse-off were raised to the level of the better-off. Since D is identical to C in equality, D should also be better than B on the above reasoning. However, Temkin believes that whether D is better than B regarding equality 'might depend on who its members were or how it came about. If B were transformed into C by raising most of the worst-off, and then D resulted from proportional increases in C's populations, then most might agree that D's inequality is better than B's. However, if D resulted from B via mere increases in the population of B's better-off group, I [i.e., Temkin] think most would rightly reject the claim that D was better than B regarding equality.'[19]

I believe that the above view may well be based on the confusion of non-ultimate considerations with ultimate values. In itself, D is clearly better than B, but the process attempting to generate D from B via

[17] For example, Locke's asymetric utilitarian solution to Parfit's population problem is open to the same objection; see D. Locke, 'The Parfit Population Problem', *Philosophy*, lxii (1987), 131–57.

[18] L. S. Temkin, 'Intransitivity and the Mere Addition Paradox', *Philosophy and Public Affairs*, xi (1987), 172.

[19] Ibid., 172–3.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

increases in the population of B's better-off group may affect the feelings of the worst-off group. Seeing that there are more people better-off than them now, they may quite likely feel worse-off. However, this consideration means that the resulting population-welfare profile is no longer D but D′ where the worse-off group has even lower welfare level. Hence, if D itself is presented for ultimate comparison with B, it must already be inclusive of *all* relevant considerations and hence better than B.

Next, consider Temkin's argument in favour of *perfectionism* (P):

> Espoused in different forms by Aristotle and Nietzsche, P has received relatively little attention from philosophers of late. Nevertheless ... it is difficult not to be strongly pulled toward some version of P as *one* ideal, among others, deserving of value. For our purposes, let us say that according to P, A is better than B if *some* of A's members are better off, or lead fuller, richer lives, than the members of B′.[20]

In other words, perfectionism maximizing the welfare of the best-off person is the exact opposite of maximin. For example, A = (1,000, 0, 0, 0, ...) where the numbers stand for the welfare levels of individuals, is regarded as a better social state than B=(999, 999, 999, 999, ...) by perfectionism or maximax.

I find maximin absurd enough, but maximax is worse. How could such absurdity appeal to reasonable men? A possible explanation is their confusion between non-ultimate considerations with ultimate values. Despite lower aggregate welfare and less equal distribution, the higher quality of life for some individuals may be desirable if this allows this fortunate group of individuals to engage in say artistic and scientific pursuits which eventually lead to the advancement of the welfare levels for most people. However, this desirability is not due to the ultimate or intrinsic values of perfectionism as such, but to its long-term utilitarian values. In discussing ultimate moral principles, we should include all relevant considerations and effects, direct or indirect, short-term or long-term, into the comparison. The effects of artistic and scientific pursuits on future welfare levels should have been included in the specification of A and B. With this correct specification, maximax should have no place at all in any reasonable person's ethics, for intuitively obvious reasons.

From Hume's law, it is impossible to disprove any basic value judgment based only on positive (as distinct from normative) arguments. Hence, even after the distinction between non-ultimate considerations and ultimate values has been explained, one may still stick to maximax. This could still be a logical position to take, even though it is one regarded by most people as extremely unreasonable. However, it would not be logical to adhere simultaneously to both maximax (or

[20] Ibid., 155.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

perfectionism) on the one hand and to maximin on the other, since these two principles are intrinsically opposite to each other in their basic morality. But, alas, this is in fact the position taken by Temkin who argues simultaneously in favour of maximin and maximax.[21] I can only explain such inconsistency by the confusion between non-ultimate considerations and basic values. The moral incompatibility between maximin and maximax is similar to the logical incompatibility between black and white. In answer to the question, 'is this cup black or white?', one may answer either black or white or neither or even some mid-way answer, e.g., grey, half-black and half-white, black when wet and white when cold, etc., but one cannot logically answer that it is all black all the time *and* all white all the time, as these two properties are mutually exclusive.

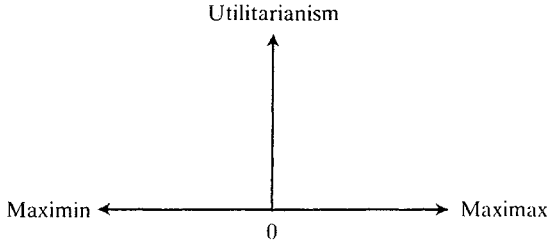

Diagram 2.

As illustrated in Diagram 2, maximin and maximax are two diametrically opposite positions, with ultilitarianism as some kind of a middle ground. One may choose any one ray from the origin or even declare something like, 'my position is somewhere between maximin and utilitarianism but I am not very certain where exactly I am located'. However, one cannot logically declare adherence to both maximin and maximax. (In fact, on top of maximin and maximax, Temkin also declares allegiance to utilitarianism and equality as desirable moral principles. In my view, equality (in the distribution of income and other primary goods) is just a way to promote aggregate welfare.[22] If one finds maximin, utilitarianism, and maximax equally attractive, one should really settle for utilitarianism as a reasonable compromise between extremes. One's original attraction to both maximin and maximax is either due to the confusion between non-ultimate considerations and ultimate values, or to inconsistent or unclear reasoning.

[21] Ibid.

[22] The preference for equality in Welfare as such may well be due to 'utility illusion', the double discounting of the incomes of the well-off; see Y. K. Ng, 'Bentham or Bergson?', pp. 545–70.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

Next, consider Dworkin's[23] argument[24] (supported by Alexander and Schwerzschild) that it is absurd to impoverish some people in order to compensate other people for the latter's extreme disappointment at the lack of life on Mars. I am not going to argue that such compensation is not absurd (in practically all cases). Rather, they (the three authors cited) use this argument in such a way that confuses non-ultimate considerations with ultimate values and hence obtain conclusions contrary to utilitarianism. I agree that the said compensation is absurd for the following reasons (all consistent with utilitarianism):

1. Such compensations would encourage unverifiable claims of extreme disappointments on many things in the future;
2. When some people suffer some drastic welfare losses through say income losses due to unexpected events beyond their control, it may be desirable for the society as a whole to compensate them, as this may increase aggregate welfare, e.g., the marginal utilities of income of the adversely affected people being increased through a reduction in income levels. This factor is unlikely to apply in the case of extreme disappointment at the lack of life on Mars.
3. For cases of unexpected income losses mentioned above, compensation may also make sense due to such factors as the promotion of social coherence through mutual concern. Due to people's perception as it is, this factor is again unlikely to be applicable for the case of disappointment over the lack of life on Mars.

If none of the above nor any other possible factors making compensation for disappointment different from compensation for unexpected income losses in their welfare-enhancing effects are applicable, in other words, if compensation for disappointment at the lack of life on Mars is welfare-enhancing taking everything into account, then I see no absurdity in compensation.

The age-long controversy regarding whether utility differs only in quantity or also in quality may also be resolved by recognizing the fallacy discussed in this section. My enjoyment of beautiful scenery is of course qualitatively different, in a certain sense, from my enjoyment of eating an apple. But these different types of enjoyment can be quantitatively compared. Here, we are concerned only with qualitative differences that cannot be reduced to some quantitative differences. In my view, there exists no such qualitative differences in utility. Poetry may yield a higher type of utility than push-pin but only in the sense that it takes more refined training for one to derive utility from the former than from the latter. For the same amount of utility, poetry utility is no more significant than push-pin utility, unless the former

[23] R. Dworkin, 'What is Equality?', *Philosophy and Public Affairs*, x (1981), 201–4.

[24] L. Alexander and M. Schwartzchild, 'Liberalism, Neutrality, and Equality of Welfare vs. Equality of Resources', *Philosophy and Public Affairs*, xvi (1987), 90.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

leads, directly or indirectly, to more utilities. But then this is due to quantitatively more utility, ultimately speaking. Thus, the preference for poetry utility over push-pin utility as such may well be based on confusing non-ultimate comparisons with basic values.

At least in some essential aspects, the fallacy of confusing non-ultimate considerations with basic values may be put in an abstract form in terms of the arguments in the social objective function. At any moment in time, our objective appears to be a function of the levels of individual welfare *and* (the degrees of observance of) some moral principles and/or rights.[25] Symbolically:

$$0_t = 0_t(w_t^1, w_t^2, \ldots, w_t^n; p_t^1, p_t^2, \ldots, p_t^m)$$

where 0 measures the degree our objective is achieved, t is the time subscript, and the superscripts stand for the numbering of individuals and moral principles, e.g., p_t^5 is the degree the fifth principle is observed at time t.[26]

If the reason why the p's enter the objective function is due to their contribution to the w's in the future, as they should be, then the ultimate or overall objective function should just be an integral of social welfare W_t as a function of individual welfares, i.e., $W = \int_0^\infty W_t (w_t^1, w_t^2 \ldots, w_t^n)\, dt$ (with possible discounting to the present-value terms if W_t in the future are not certain). It should no longer have the p's in its argument, simply the w's for all times. The confusion of non-ultimate considerations with basic values is not to recognize the above distinction and use 0_t instead of W for ultimate comparison and hence mistakenly regard the p's as of ultimate significance.

V. THE DISTINCTION BETWEEN IDEAL MORALITY AND SELF-INTEREST

Another confusion that adversely affects people's support for utilitarianism is the insufficient distinction between ideal morality and self-interest. This confusion is often made by utilitarians themselves.

Ignoring imperfect knowledge and irrational preferences, the preference of an individual i may be represented by the following utility function:

$$u^i = u^i (w^i; w^1, w^2, \ldots, w^{i-1}, w^{i+1}, \ldots, w^n)$$

[25] Some people may take the (I think naive) position that a right or principle is either fully observed or not at all. In that case, the degree of observance assumes a value of either one or zero.

[26] For simplicity, the possible dependence of 0_t on variables at times other than t is ignored.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

This function allows the individual to take account of his own welfare w^i as well as the welfares of other individuals. For simplicity, further ignore his differentiation between other individuals, and writing $v^i \equiv w^1 + w^2 + \ldots, + w^{i-1} + w^{i+1} + \ldots, + w^n$, (i.e., v^i is the sum of all individual welfares other than that of individual i himself),[27] we have

$$u^i = u^i(w^i, v^i)$$

with a specific summation form as $u^i = w^i + \alpha v^i$.

With this utility function, ideal utilitarian morality requires $\alpha = 1$ while the pursuit of pure self-interest results in $\alpha = 0$. In *specific* acts of heroism or altruism, some individuals may behave or choose such as to display an α larger than one. Some moral philosophers define altruism as $\alpha > 1$ but I regard this as ultra-altruism. A positive α is enough for *some* degree of altruism. In comparison to perfect altruism ($\alpha = 1$), ultra-altruism results in reduced aggregate welfare, except for its possible extra positive effect on the encouragement of further altruistic behaviour.

For behaviour or choice over a long period, I doubt that anyone has a value of α averaging to equal one. This ideal degree of morality is unlikely to have been attained even by saints, religious leaders and fervent followers.

Acts that appear very altruistic may be purely self-interested since the actor may do them because they increase their own welfare. Someone with $\alpha = 0$ may do certain things benefiting others because doing these apparently altruistic acts makes him feel better. The apparently altruistic act A enters his utility function only through his welfare function but not separately, i.e., $u^i = w^i(A, \ldots) + 0v^i$ where the zero before v^i indicates his absence of any concern for the welfare of others. It is of course also possible that someone may do certain altruistic acts because of his concern for the welfare of others ($\alpha > 0$) but then the doing of this act also makes him feel better.

Ignoring imperfect knowledge and irrational preferences, we may write the utility function of individual i as,

$$u^i = u^i\{w^i(a_1, \ldots, a_r, w^1, \ldots, w^{i-1}, w^{i+1}, \ldots, w^n), w^1, \ldots, w^{i-1}, w^{i+1}, \ldots, w^n)\}$$

where the a's represent some objective facts, e.g., amounts of goods consumed, actions taken, by i or other individuals. The point to be made here is that w^j ($j \neq i$) may enter u^i through w^i or directly or both.

[27] If the individual concerned does not like this sum-ranking method, we may just define $v^i = v^i(w^1, w^2, \ldots, w^{i-1}, w^{i+1} \ldots, w^n)$ which is whatever function he uses for the welfares of other individuals.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

While the positive effect of other individual welfares on one's own welfare, i.e., $\partial w^i/\partial w^j$ $(j \neq i) > 0$, shows *sympathy*, the direct positive effect (other than through one's own welfare) of other individual welfares on one's utility, i.e., $\partial u^i/\partial w^j$ $(j \neq i) > 0$ (independent of $\partial w^i/\partial w^j$), shows *concern*. Admittedly sympathy and concern are usually interwined but they are analytically distinct and one may act to benefit others due either mainly to sympathy or mainly to concern.

Most utilitarian philosophers seem to believe that utiltarianism *requires* one to have $\alpha = 1$ (i.e., equal concern for any other individual welfare as one's own welfare). This is of course the ideal morality one may be hoping for but it is an ideal that can hardly be reached. Requiring utilitarians to have such perfect morality ($\alpha = 1$) scares people away.

A person may be a perfect utilitarian in believing that, *for the society*, we should have $w = w^1 + w^2 + \ldots . w^n$. However, for this personal behaviour, he may be quite self-concerning in having an α close to zero. This is a perfectly logical position to take.

An area in which the insufficient distinction between ideal morality and self-interest has caused unnecessary controversies is normative population theory. Among utilitarians, there are those in favour of maximizing average utility and those in favour of maximizing total or aggregate utility (i.e., average utility times the number of population). For the viewpoint of ideal morality, average utility maximization is obviously unacceptable as it would prefer 10 persons with an average utility of 100 to 1,000,000,000 persons with an average utility of 99. Moreover, it violates the *Mere Addition Principle* that adding more happy individuals without reducing the welfare of any existing individual is obviously a good thing.[28] On the other hand, the maximization of total utility leads to what many people regard as the *Repugnant Conclusion*: For any given situation A with a given population size (e.g., ten billion people), all with very high welfare (e.g., one billion utils each), there exists (hypothetically) another situation B with a much larger population (e.g., a hundred billion billion people) whose members all have lives barely worth living (e.g., one util each); yet B is preferred to A according to the principle of total utility.[29]

Many philosophers opt to get around the Repugnant Conclusion by denying the importance of prospective individuals (yet unborn individuals that would be born under some alternative). For example, Narveson[30] believes that we 'are in favour of making people happy, but neutral about making happy people'. However, in my view, an unborn person is similar to a person with zero utility; increasing the utility of

[28] J. Narveson, 'Moral Problems of Population', *The Monist*, lvii (1973), 80.
[29] D. Parfit, *Reasons and Persons*, Oxford, 1984.
[30] Narveson, p. 80.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

an existing person with zero utility to x utils is equivalent to having an extra person with x utils provided that no other person is affected in both cases. Moreover, ignoring prospective individuals will lead to the *Extinction Paradox*: Provided everyone in existence is made slightly happier, we would choose to let the human race become extinct.

In addition, being neutral about making happy people does not really avoid the Repugnant Conclusion, if we accept the compelling *Non-Antiegalitarianism*: If alternative B has the same set of individuals as in alternative A, with all individuals in B enjoying the same level of utility as each other, and with a higher total utility than A, then, other things being equal, alternative B must be regarded as better than alternative A.

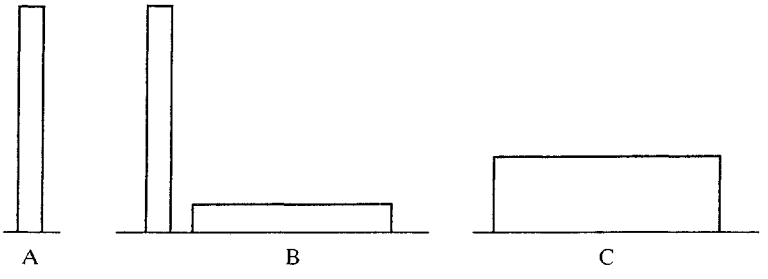

Diagram 3.

Consider Diagram 3. Neutrality about making happy people requires that we must be indifferent between B and A. Non-Antiegalitarianism requires that we prefer C to B. So from transitivity, we should prefer C to A. Either by repeated application of the above or by making the second set of individuals in B extremely large, we still obtain the Repugnant Conclusion.

Some people may find B inferior to A because the inequality in welfare levels may make the original set of individuals (those already existing in A) worse-off. This reasoning is based on the fallacy discussed in Section IV. If the alleged effect is there, the original set of individuals would no longer enjoy an unchanged welfare level in B (as in A), unless the alleged negative effect is exactly offset by other positive effects in B. In the latter case, since full compensation is achieved, there is no acceptable ground for regarding B as inferior to A, since the new set of individuals are happy.

The dilemma of average utility (with its violation of the Mere Addition Principle) versus total utility (with the Repugnant Conclusion) can easily be solved by taking the position that, in terms of ideal

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650

morality, total utility is the right principle,[31] however, where our (existing people's) self-interest may have to be sacrificed, we may not follow the rule of ideal morality but may choose to be partial to our own self-interest, with only some concern to the welfare of prospective individuals. This position for existing people (as a group) is similar to the position by an individual with respect to his treatment between his own welfare (w^i) and that of others (v^1) discussed above. For choices affecting only the welfare and the number of people in the far future (who are all unborn now), there is no reason to be partial and we should choose according to the principle of total utility maximization.[32]

VI. CONCLUDING REMARKS

From the above discussion, it is clear that, provided that non-ultimate considerations are not confused with basic values, that all relevant effects are taken into account, utilitarianism or at least welfarism is ethically compelling. However, utilitarianism seems very unpopular these days. While there have been some good utilitarian writings from time to time,[33] they are rather overwhelmed by non-utilitarian and even anti-utilitarian publications. The situation has been so lopsided that a commentator felt confident to remark that no one is a utilitarian these days. Nevertheless, judging the moral inclination towards or against utilitarianism by the standpoint of published writings could be quite misleading. Apart from pseudo anti-utilitarianism due to the confusion of non-ultimate considerations with basic values discussed above, there is the bias in favour of publishing something novel. Utilitarianism has been expounded and defended for hundreds of years so that it is extremely difficult to be novel in taking a utilitarian position. On the other hand, it is much easier to be novel in taking a non-utilitarian position or one critical of utilitarianism or welfarism. However, I find all these positions ethically unacceptable upon closer examination.

[31] For the argument that, in this perspective, the Repugnant Conclusion is not repugnant at all and that Parfit's ideal theory X either does not exist or violates the compelling Non-Antiegalitarianism, see Y. K. Ng, 'What Should we do About Future Generations?', *Economics and Philosophy*, v (1989), 235–53.

[32] Ng, 'What Should we do About Future Generations?'

[33] See J. C. Harsanyi, 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility', J. J. C. Smart and B. Williams, *Utilitarianism: For and Against*, Cambridge, 1973; R. M. Hare, 'Ethical Theory and Utilitarianism', in H. D. Lewis, *Contemporary British Philosophy*, London 1976; and J. Riley, *Liberal Utilitarianism: Social Choice Theory and J. S. Mill's Philosophy*, Cambridge, 1988.

Downloaded from http:/www.cambridge.org/core. Oxford University Museum of Natural History, on 14 Nov 2016 at 20:50:36, subject to the Cambridge Core terms of use, available at http:/www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0953820800000650