

## Utility, informed preference, or happiness: Following Harsanyi's argument to its logical conclusion

Yew-Kwang Ng<sup>1,2</sup>

<sup>1</sup> Monash University, Department of Economics, Clayton, Victoria 3168, Australia  
(e-mail: kwang.ng@buseco.monash.edu.au)

<sup>2</sup> National Taiwan University, Department of Economics, 21 Hsueh Road,  
Taipei, Taiwan

Received: 29 May 1997 / Accepted: 3 November 1997

**Abstract.** Harsanyi (1997) argues that, for normative issues, informed preferences should be used, instead of actual preferences or happiness (or welfare). Following his argument allowing him to move from actual to informed preferences to its logical conclusion forces us to use happiness instead. Where informed preferences differ from happiness due to a pure concern for the welfare of others, using the former involves multiple counting. This “concerning effect” (non-affective altruism) differs from and could be on top of the “minding effect” (affective altruism) of being happy seeing or helping others to be happy. The concerning/minding effect should be excluded/included in social decision. Non-affective altruism is shown to exist in a compelling hypothetical example. Just as actual preferences should be discounted due to the effects of ignorance and spurious preferences, informed preferences should also be discounted due to some inborn or acquired tendencies to be irrational, such as placing insufficient weights on the welfare of the future, maximizing our biological fitness instead of our welfare. Harsanyi's old result on utilitarianism is however defended against criticisms in the last decade.

Harsanyi (1997) argues, among other things, that in welfare economics and ethics, what are important are people's *informed preferences*, rather than either their *actual preferences* (as emphasized by modern economists) or their *happiness* (as emphasized by early utilitarians). The main purpose of this paper is to argue that, pursuing Harsanyi's argument that allows him to move from actual to informed preferences to its logical conclusion forces us

to happiness as the ultimately important thing. The early utilitarians were right after all! Since I personally approve of Harsanyi's basic argument, I regard myself as his follower who becomes more Catholic than the Pope. (It is not denied that, in practice, the practical difficulties and undesirable side-effects of the *procedure* of using happiness instead of preferences have to be taken into account. Thus, even if we ultimately wish to maximize the aggregate happiness of people, it may be best in practice to maximize their aggregate preferences in most instances. This important consideration will be largely ignored in this paper.) The secondary objective is to give a brief defence of Harsanyi's (1953, 1955) much earlier argument for utilitarianism (social welfare as a sum of individual utilities) that has received some criticisms in the last decade. The argument (e.g. Roemer 1996) that Harsanyi's result is irrelevant to utilitarianism is based on the point that the VNM (von Neumann-Morgenstern) utility is unrelated to the subjective and interpersonally comparable cardinal utility needed for a social welfare function. Harsanyi's position is defended by showing that the two types of utility are the same (apart from an indeterminate zero point for the former that is irrelevant for utilitarianism concerning the same set of people).

### 1. Non-affective altruism: The pure concern for the welfare of others

A person's actual preferences are "indicated by his choice behavior and by his verbal statements". His informed preferences are "defined as the *hypothetical* preferences he *would have* if he had all the relevant information and had made full use of this information" (Harsanyi 1997, p. 133). In addition, to qualify as informed preferences, "our preferences should be *genuine* preferences rather than *spurious* ones" (p. 134). There are various classes of spurious preferences including *compulsive behavior* and *self-deception*.

Now, let us compare Harsanyi's distinction outlined above with my distinction between happiness (or welfare, defined to be the same) and (actual) preference (Ng 1979, 1983). (From now, "actual" will be dropped unless when emphasis is needed while "informed" will be kept throughout.) The preference of an individual may differ from his welfare due to either one or more of the following exhaustive reasons: 1. ignorance or imperfect foresight, 2. a concern for the welfare of others (the definition of "others" could be broad enough to include animal welfare), and 3. irrational preferences. The last is defined to be any divergence from welfare other than due to the first two factors. (This definition makes the tripartite classification exhaustive, though some readers may query the use of terminology.)

Comparing the above two schemes of distinction, it is clear that there should be a complete agreement where the divergence is due to ignorance, imperfect foresight, or misinformation. If a person's preference of medical treatment A over B is based on misinformation, the actual superiority of B over A seems uncontroversial. For this type of divergence, economists' preference for using actual preferences may still be sustained on the practical grounds of: 1. the difficulties of discovering the informed preferences; 2. the

possible unfavourable side-effects of diverging from the actual preferences of people. Where these two considerations are not important and/or the degree of ignorance is more important, then it is uncontroversial that welfare or informed preferences should prevail over actual preferences, as witnessed by the prevalence and widespread support of fluoridation of water and prohibition of hard drugs.

To a large extent, the spurious preferences of Harsanyi corresponds with my concept of irrationality. (I will return to discuss some possible differences later.) Thus, whether informed preference or welfare should be used for social evaluation depends mainly on how the divergence between preference (actual or informed) and welfare due to a concern for the welfare of others (non-affective altruism/malice) should be dealt with. I wish to strongly argue that such a concern should be ignored in social evaluation. In other words, social evaluation should take account of happiness or individual welfare rather than preferences (actual or informed). This is so for the simple reason that otherwise double or rather multiple counting will be involved. (Harsanyi allows for “truly altruistic actions” which seems to correspond to my “concern for the welfare of others” or non-affective altruism; however, he does not discuss the implication of this for the choice of individual welfares vs. utilities as the appropriate arguments in the social welfare function. Elsewhere, Harsanyi and others appear to reject all forms of altruism from consideration in social choice on the ground of multiple counting. As argued below, *affective* altruism should not be so excluded.)

For simplicity, consider a society of two individuals. Individual 1 is self-concerning and maximizes his own welfare, with utility function  $u^1 = w^1$ . Individual 2 has a substantial concern for the welfare of individual 1, with utility function  $u^2 = w^2 + 0.5 w^1$ . Also for simplicity (but not essential for the argument), suppose that the relevant social welfare function is utilitarian, maximizing the unweighted sum of individual values. The question here is whether the individual values should be individual utilities or individual welfares. Note also that, for this exercise (i.e. summing, and in fact any other reasonable method of aggregating, individual utilities or welfares to arrive at social welfare) to be possible, we must have a framework of interpersonally comparable cardinal individual utilities or welfares. (The necessity of interpersonal comparison is established by Sen 1970; the necessity of cardinal utility or the insufficiency of ordinalism even with a fixed set of individual preferences is established by Kemp and Ng 1976, and Parks 1976. On a method actually used to measure interpersonally comparable cardinal utilities, see Ng 1996a.) Thus, in the following table, the welfare and utility figures are cardinal and interpersonally comparable.

Social states	$w^1$	$w^2$	$u^1$	$u^2$	$\Sigma w^i$	$\Sigma u^i$
$x$	2	7	2	8	9	10
$y$	6	2	6	5	8	11
$z$	8	5	8	9	13	17

For the situation depicted, should the society choose  $x$  or  $y$  for the choice between  $x$  and  $y$  only (with  $z$  not feasible)? If  $\Sigma w^i$  is used,  $x$  is preferred to  $y$ ; if  $\Sigma u^i$  is used,  $y$  is preferred to  $x$ . In my view, it is clear that  $\Sigma w^i$  should be used and  $x$  should be socially preferred to  $y$ . In fact, since the individual welfare profile of  $x$  is (2, 7) while that of  $y$  is (6, 2), in terms of individual welfares,  $x$  will be preferred by any social welfare function that is anonymous and increasing in individual welfares. Though  $x$  has a lower  $\Sigma u^i$ , this is so only because the substantial concern of individual 2 for the welfare of individual 1. If individual 2 were to make the choice herself, she would prefer  $z$  (if feasible) to  $x$ . Though her welfare is lower at  $x$  than at  $z$ , her concern for the welfare of individual 1 more than offsets this as  $w^1$  is much higher at  $z$  than at  $x$ . However, her concern for the welfare of individual 2 is not sufficient for her to prefer  $x$  to  $y$ . For social choice, the use of  $\Sigma w^i$  already takes full account of  $w^1$ , i.e.  $w^1$  is already treated at a par with  $w^2$ . Thus, there is no further need to take account of the concern of individual 2 for  $w^1$ . To do so would involve the double counting (multiple counting in the case of many non-self-concerning individuals) of the welfare of individual 1. In using  $\Sigma w^i$ ,  $w^1$  is counted fully (i.e. with the weight of unity) under  $u^1$ , and then counted again at the weight of 0.5 under  $u^2$ . It is thus counted 1.5 times, while  $w^2$  is counted only once.

It may be thought that, since  $w^1$  enters  $u^2$  at the weight of 0.5 while  $w^2$  does not enter  $u^1$ , it may be right to count  $w^1$  1.5 times while counting  $w^2$  only once. This is inappropriate. Consider the concrete example of parental choice. Usually, parents have sufficient concern for the welfare of their children such that the society does not find it necessary to interfere and just let parents choose for the whole family. (There are also reasons of the practical difficulties and side effects of interference which we shall ignore for simplicity.) However, to tackle some special cases of gross parental negligence, there are appropriate legislations. Nevertheless, even ideally, such legislation should only aim to ensure that the welfare of children are taken fully (i.e. at a par with that of the parents) into account, not that the welfare of children should be more important than that of the parents. To do the latter would be making the mistake of correcting a mistake excessively.

A reason why some people may prefer using  $\Sigma u^i$  rather than  $\Sigma w^i$  (or other functions of individual utilities rather than individual welfares) is due to a failure to distinguish “minding” from “concerning”. A self-minding person is one whose welfare (or happiness) is not affected by the welfare levels of others. A person may feel bad knowing that there are people suffering in Africa. This is a “minding” effect or affective altruism. A self-concerning person is one whose utility (or preference) is not affected by the welfare levels of others, except possibly through their effects on his/her own welfare (i.e. through the “minding” effect). A person may choose  $x$  over  $y$  even if her welfare is lower at  $x$  than at  $y$  because the welfare levels of others are higher at  $y$  than at  $x$ . This is a “concerning” effect or non-affective altruism. A self-centring person is one who is both self-minding and self-concerning.

In functional forms and ignoring ignorance, we may write the utility of a rational person as

$$u^i = u^i(w^1, w^2, \dots, w^I)$$

where  $u$  = utility,  $w$  = welfare, and a superscript indicates the person concerned, and  $I$  = the number of individuals concerned. And,

$$w^i = w^i(a, b, c, \dots, w^1, w^2, \dots, w^{i-1}, w^{i+1}, \dots, w^I)$$

where  $a, b, c, \dots$  are some variables like consumption that enter the welfare function but are not the focus here. The “minding” effect is  $\partial w^i / \partial w^j, j \neq i$  and the “concerning” effect is the direct effect of  $w^j$  on  $u^i$  without going through  $\partial w^i / \partial w^j$ .

It is true that these two effects are usually intertwined. If a person is concerned with the welfare of another person, she is also likely to be non-self-minding towards his welfare. (The reverse may also be true though with lower force, as the minding effect is usually stronger and more prevalent than the concerning effect.) Thus, parents are typically happy knowing that their children are happy (minding effect) *and* also willing to choose something against their welfare for the welfare of their children (concerning effect). It is true that they will feel good knowing that they are sacrificing for their children and that the welfare of their children increases, but if this is not sufficient to offset their welfare loss due to other factors (such as a lower level of consumption, less leisure, worse health), the concerning effect is involved over and above the minding effect.

As the minding effect actually affects the welfare level, it is already taken fully into account in the social decision in accordance to  $\Sigma w^i$ , as it should be. For the concerning effect, the welfare of the concerning person is not affected, but her preference is affected by her concern for the welfare of others. However, in using say the unweighted sum of individual values, the welfare of others is already taken fully into account, i.e. is already fully “concerned” with. Hence, there is no further need to take the concern of this person for the welfare of others into account.

The existence of non-affective altruism (the concerning effect) may be doubted (especially by economists, as mentioned to me on several occasions). It may be argued that, if Ms. 2 is willing to choose  $z$  over  $x$  in consideration of the welfare of Mr. 1, she must be happier with  $z$  than with  $x$ . So,  $u^i$  and  $w^i$  must always go together such that the situation depicted in the table above where  $u^2$  and  $w^2$  conflict each other with respect to the pair  $(x, z)$  cannot logically arise. In other words, only “minding” is possible; “concerning” is not possible. This argument is incorrect. As conceded above, the two effects usually intertwine with each other, and “minding” is more prevalent than “concerning”. Many apparently “concerning” effects may actually be “minding” effects upon closer examination. However, some truly pure “concerning” effects are possible. To see this, consider the following purely hypothetical construction which nevertheless illustrates the point most dramatically.

Suppose that you are commanded by the all-powerful Devil to press either button A or B within 2 seconds. You know with certainty that the following outcomes will happen depending on which button you press.

*A:* You will go to “Bliss” with your happiness level at 1,000 trillion units. (Those who prefer concreteness may imagine a health-enhancing island with all the material supplies you want plus 100 abiding partners of your dream.) All others on earth will go to hell with happiness level at *minus* 1,000 trillion units each.

*B:* You will go to “Bliss Minus” with your happiness level at 999 trillion units. (The same as above but with 99 partners.) All others will go to “Niceland” with happiness level at 100 trillion units each.

*C:* If you do not press any button within the 2 seconds, you *and* all others will go to hell.

There will be no communication between Bliss, Niceland, hell, etc. You will lose all memory of the present world once you press either button or fail to press within the 2 seconds. So you will not have any guilt feelings in Bliss or Bliss Minus. Within the 2 seconds, you will be too pre-occupied with pressing the right button that your welfare will be zero whichever button you press. (It is too brief to experience any significant amount of welfare anyway.)

Thus, by construction, your welfare will be higher with A than with B. However, most people would choose B. (I would have not the slightest hesitation at all in choosing B.) The choice of B over A exhibits non-affective altruism, though it may not be a very strong one. If you still prefer A, change the happiness level of Bliss Minus into 999.999999 trillion units. If you still prefer A, then you may really be perfectly self-concerning. But how could you condemn all others to hell for a tiny fractional increase in your own welfare?

Now consider more realistic choices. It is true that, as parents, we usually feel happy doing something for our children (or other loved ones) like sacrificing our time, effort, money. Thus, the importance of the “minding” effect is uncontroversial. However, are we willing to make such sacrifices *only if* our loss in welfare due to the sacrificed time, effort, or money is more than made up by our warm-glow feeling of helping our children? If the loss is not fully made up but yet the welfare of our children will be significantly increased at a moderate or even small net loss to our own welfare, will not at least some of us be willing to do that?

If we have a true concern (over and above the “minding” effect) for the welfare of our children, can we not have a similar (though lesser in degree) concern for our siblings, our relatives, our friends, our fellow countrymen/women, the whole humankind, and eventually all sentient? (For some evidence for true altruism, see Hoffman 1981; Monroe 1996.)

It is well known that animal behaviour is largely determined by genetic programming to maximize the inclusive fitness (which consists of the indi-

vidual's own fitness as well as the effects on the fitness of genetically related neighbors, weighted by the degree of relatedness; see, Hamilton 1964; Wilson 1975). We care about our children/relatives because they carry/share our genes. Sometimes one (or one's spouse) may wonder why one is so willing to help one's siblings (or other relatives) whom one does not really love much, not knowing that perhaps the genetically determined subconscious inclination may play a role. Thus, the existence of some non-self-centring effects is a biological necessity. However, whether the biologically determined non-self-centring effects manifests mainly/only in the "minding" or the "concerning" effects remains to be explored. Of course, we are much more influenced by nurture than other animals and also have a higher, if not exclusive, sense of morality. It may be conjectured that, the more important is the influence of nurture and the more important is the sense of morality, the more likely is it for the true concerning effect to be present and significant.

From the above discussion, it may be concluded that true concern (i.e. on top of or over and above the "minding" effect) for the welfare of others does exist. Moreover, when preferences (actual or informed) differ from welfare due to a concern for the welfare of others, it should be the individual welfare rather than the utility values that enter the social welfare function.

While Harsanyi does not discuss the appropriate treatment of altruistic preferences in his 1997 paper, he did discuss it elsewhere (Harsanyi 1995, p. 325). He suggested that all "external" preferences, even if altruistic, should be excluded in social welfare consideration, presumably based on similar reasoning as our argument above. Here, external preferences are "preferences for assignment of goods and opportunities to others" (Dworkin 1977). The belief in the need to exclude altruism (or its opposite, malice) is held by a number of other authors discussing the issue (e.g. Hammond 1987/1995). My point here is that these authors do not distinguish between affective (i.e. the minding effect) and non-affective (the concerning effect) altruism and seem to suggest that they should all be disregarded. As argued above, while the concerning effect should be excluded to avoid multiple counting, the minding effect should be included.

To see the reasonableness of including the minding effect, consider the hypothetical example that either S or M has to die and a lot is drawn to decide who has to die. Suppose that the two are equally old, capable, etc., except that S is single with no close friends and M has many close relatives who love her very much and will suffer a lot from her death. Then, other things being equal, most people will hope that the lot will turn out to let M survive rather than the other way round. The death of M will cause more suffering. Counting this suffering does not involve double counting. The minding effect should be treated differently from the concerning effect. If I feel sad or happy for certain event, such feelings should certainly count in assessing the social desirability of that event. On the other hand, if I have some non-affective concern for the welfare of others, such a pure concern should not be counted if the welfare of these others are already fully concerned for in the social welfare function.

How could we say that the sorrow of a mother from the death of her child should be disregarded as it is “external” preference? It could cause more suffering on her than starvation! Perhaps Dworkin would reply that the death of a child would adversely affect the opportunities of the mother, and hence the mother’s preference here is not wholly external. However, for any definition of opportunities (unless it is defined to coincide with welfare), one can revise the example such that the preference is external as defined by Dworkin (i.e. regarding the assignment of goods and opportunities to others) and yet the person concerned (who has the external preferences) genuinely suffers a lot subjectively. Such external preferences should not be ignored. (However, we may have to disregard many external preferences even when they genuinely affect people’s welfare, on grounds of practical difficulties and undesirable side effects; but this problem has been abstracted away in this paper.)

## 2. Irrational preferences

As mentioned above, Harsanyi regards “spurious” preferences as not real preferences and hence should not be used for normative purposes. He gives two examples of spurious preferences. One is “compulsive behavior” where, for example, “some neurotics wash their hands far too many times a day for no obvious reason. Their behavior may be to some extent voluntary and to that extent a result of their own *preferences* to act in this way. But at a deeper level, it is obviously a result of a more or less irresistible *inner compulsion*, very much *contrary* to their true preferences” (p. 135). This may be regarded as some degree of psychological sickness due perhaps to some subconscious need to wash away some guilt or “dirtiness”. Such spurious preferences should not be treated on a par as normal preferences. Thus, curing such sickness should be regarded as an improvement but we do not want to change a person’s normal preference of say apples over pears. However, before the sickness is cured, we may still want to “respect” his preference as, for example, denying him the water to wash his hands may make him feel very uncomfortable. But why is curing his compulsive behaviour good while denying him water bad? In my view, this is due to different effects on his (and perhaps also on others’) happiness. Thus, again, happiness is more fundamental than preferences.

The second example of spurious preferences is “self-deception”. “Some people pretend to have, and may in the end even convince themselves that they have, some preferences they think to be fashionable and sophisticated – even though their real preferences may be quite different, or even though after a point they may not really know themselves what their real preferences are. Devotees of various esoteric art forms of questionable aesthetic value often form coteries that seem to display this kind of behavior” (Harsanyi 1997, p. 135). I do not know much about these esoteric art forms but it could be debatable to call the preferences of their devotees self-deception.

Harsanyi believes that informed preferences, by definition, will always be based on all the relevant information, so that they will always be in agreement with our real interests. However, it is debatable that people's preferences must always coincide with their interests if they are fully informed. The meaning of "interests" here is vague. If it means whatever people prefer under full information, then it coincides with informed preference tautologically but the terminology of "interest" is misleading. If it means welfare, happiness, or something similar, then it is not true that people's preferences when fully informed always coincide with their interest, for at least two reasons. First, there is non-affective altruism (and possibly also malice), as discussed in the last section. Secondly, biological and psychological motivational factors may affect preference or behaviour in ways not fully consistent with welfare, as argued below.

There are a number of causes that may make preferences differ from happiness other than ignorance and a concern for the welfare of others, hence irrational according to my definition. The following two (may not be completely independent) causes may both be explained, at least partly, by some biological factors. (There are also some psychological conditioning effects that cause irrational behavior briefly touched on by Harsanyi. For some other causes of irrational preferences, see Ng 1989a.)

First, there is the tendency of many people to discount the future too much or even to ignore it completely. This is widely noted, including by economists. For example, Pigou (1929, p. 25) called it the "faulty telescopic faculty", Ramsey (1928, p. 543) called it "weakness of imagination" about the future, Harrod (1948, p. 40) regarded it as the "conquest of reason by passion". A discount on future consumption, income, and any other monetary values is rational as a dollar now can be transformed into more than a dollar in the future. A discount on future utility may still be rational if the realization of the future utility is uncertain. (For healthy people, this uncertainty is usually very small.) Discounting the future for more than these acceptable reasons is probably irrational. A manifestation of this irrationality is the insufficient amount of savings for old age, necessitating compulsory and heavily subsidized superannuation schemes. I came across an extreme example of such under-saving during a survey regarding how much people would be willing to save more if the rate of interest were higher (Ng 1992). The question implicitly assumed that everyone did some saving, as the answers were in terms of how many percentages more one would save. One subject declared that he did not save anything. I then asked him to change the answers to be chosen from "saving 20% more" into "saving \$20 more per month", etc. He still said that he could not be induced to save anything even at annual interest rates of hundreds of percent. It is only when I said, "If a dollar saved now could become a million dollar next year, would you save?" that he admitted he would save then. I was careful enough to find out that this healthy-looking young man was not expecting early death from a terminating disease or the like.

The behavior of most other animals is largely determined by pre-programmed instincts rather than the careful calculation of the present costs versus the future benefits. The storing of food by ants, the bury of nuts by squirrels, etc. are largely, if not completely, instinctive. If calculated choices are made by animals, they are largely confined to sizing up the current situation to decide the best move at the moment, like fight or flight. The ability to anticipate the rewards in the fairly distant future requires much more “reason”, “imagination”, and “telescopic faculty” than normally cost-effective to program in most other species. However, we know that we are endowed with some such faculty. (This is not to say that we are not also partly driven by what Keynes called “the animal spirit”, including the accumulation instinct.) Nevertheless, since this advanced faculty is almost completely absent in most other species, it is natural to expect that it is not fully developed even in our own species. Moreover, different members of our species may be endowed with different degrees of such faculty. The existence of a significant proportion of our species who do not possess a full telescopic faculty is thus not surprising.

Secondly, there are the excessive temptation of pleasure (especially present pleasure vs. future costs, hence related to the preceding cause) and the powerful biological drives. After the evolution of flexible species (defined as one the behavior of whose members is not completely determined by the automatic programmed responses but also by choice), natural selection ensured that the flexible choices made were consistent with fitness by endowing the flexible species with the reward-penalty system. Thus, eating when hungry and mating with fertile members of the opposite sex are rewarded with pleasure and damages to the body are penalized with pain. (This makes the flexible species also “rational” as defined in Ng 1996b which shows that complex niches favour rational species which make the environment more complex, leading to a virtuous cycle that accelerated the rate of evolution, partly explaining the dramatic speed of evolution based mainly on random mutation and natural selection, a speed doubted by creationists.) On top of the *ex-post* rewards and penalties, we are also endowed with inner drives to satisfy the fitness-enhancing functions like mating. On the whole, these powerful temptations and drives works in the right direction, making us do things that both enhance our biological fitness and psychological welfare. However, since evolution is largely fitness-maximization and the welfare-enhancing aspect is only indirectly to enhance fitness, some divergence between our behavior and our welfare is unavoidable, as our behavior is not completely determined by rational calculation but also partly by the programmed inclination, including the drives. (See Ng 1995 on the divergence between fitness and welfare maximization especially with respect to the number of offsprings.) As an example, adolescent girls and boys often engaged in careless sexual acts propelled by their sexual drive and tempted by the sexual pleasure even at high risk to their long-term welfare, such as having unwanted pregnancies and contracting aids. While this is partly due to ignorance, the role of biological drives cannot be denied.

Consider a specific example. Suppose that a person agrees that, for choices involving risks, the correct thing to do is to maximize expected welfare (assuming no effects on the welfare of others) and also actually do so for most choices. However, for choices concerning seeking sexual activities, he chooses  $x$  over  $y$  though his expected welfare is lower with  $x$  than with  $y$  and that he knows this to be the case. Here,  $x$  may involve having sex with many persons without clear knowledge (this knowledge is assumed to be not feasible to obtain and hence not relevant) about their infectability of aids. His (expected) welfare-reducing choice of  $x$  may be due to the biological inclination to seek many sexual encounters. He knows that doing so has a non-insignificant chance of contracting aids and hence is welfare-reducing. He has all the relevant feasible information and yet chooses (due to the powerful sex drive)  $x$  that he knows to be of lower expected welfare. (This is not really a hypothetical example. I am confident that, out of 100 average adult males, at least 10 have actually made such choices. If one wants more solid evidence, one may look at the frequency of prostitution and extra-marital sex.) Should we call this preference informed as the person has all the relevant feasible information or not informed because it is not in agreement with his real interests?

The above two causes of irrational preference illustrate the point that, due either to imperfection in our endowed faculty or the biological bias in favour of reproductive fitness, we may do things not quite consistent with our welfare. The issue here is that, for normative purposes, should we use welfare or actual preferences/behavior. Clearly, we should use welfare instead of behaviour dictated by biological fitness. An old Chinese dictum says, "Out of the three unfilial acts, not having offspring is the greatest". However, for the human species as a whole, we are certainly not getting smaller in population size. Moreover, a long-run social welfare function accounting for the welfare of future generations should account for that. If we go for biological fitness, we will prefer unlimited procreation even if that means that we will all be suffering to a smaller population with a higher aggregate welfare. "*We*" are the feeling selves that cares ultimately about our welfare (positive minus negative affective feelings). *We* are not them, the unfeeling genes that, through random mutation and natural selection, programmed us to maximize fitness. Unlike other species who are almost completely controlled by their genes and the environment, we have learned to change our fate by using such measures as birth controls. For normative issues, it is *our* welfare, rather than the selected random dictates of the unfeeling genes, that should count.

For those who agree with Harsanyi in rejecting spurious preferences (including preferences for "various esoteric art forms of questionable aesthetic value") for normative purposes, it seems likely that irrational preferences due to biological drives and imperfections and psychological conditioning should similarly be rejected. However, there is a class of preferences ("*autonomous* desire *not* based on hedonistic considerations") that may be classified as irrational using my definition but insisted by Harsanyi to be respectable, as discussed in the next section.

### 3. Autonomous desires

Instances of autonomous desires not based on hedonistic considerations given by Harsanyi (p. 132–3) includes: altruistic desires, desires for accomplishments, our natural desire to satisfy our curiosity. I allow for altruistic desires to be rational as a concern for the welfare of others. This has been discussed above.

I quite agree that some of our desires are not based on hedonistic considerations. However, I believe that the satisfaction of desires or preferences as such has no intrinsic normative value; it is the effects on happiness that is ultimately valuable. Otherwise, why are spurious preferences not normatively important?

We do have a *natural* desire to satisfy our curiosity. We are probably both born and brought up to be so. As our species manages successfully to survive mainly based on its superior intelligence and knowledge, curiosity has a fitness-enhancing effect. Hence, we are also rewarded in satisfying our curiosity by feeling very good. And the increased knowledge contributes to future success. Thus, the satisfaction of this desire is generally consistent with our welfare. However, there are cases where the satisfaction of our curiosity actually makes us worse off. If a person see a box in her office, she will naturally open it to have a look. If a colleague prevents her from doing so by seizing the box away from her, she will be made a little unhappy. However, if the box contains a poisonous snake, she will be very grateful to her colleague. It may be said that, in this case, it is only the uninformed preference that is violated, not the informed one. But suppose the box contains a photograph of her mother being raped by a soldier . Then, even after being informed of that, she may still be unable to overcome her curiosity and will open it to have a look. If this disturbs her a lot, she will be made much more worse off than by the seizing away of the box from her. Suppose that there are no benefits, directly or indirectly, of so looking at the photograph, to her or to others and that there are no side effects of seizing the box away from her. Most people will agree that, in this case, seizing away the box from her is the right thing to do. If I were her, I would be most grateful for that. The satisfaction of my (even if informed) preferences as such has no normative significance for me; it is important only because, in most cases, it makes me (and/or others) happier, directly or indirectly.

Now consider the desire for accomplishment. This may also have some genetic basis and may be related to the accumulation instinct possessed by many animal species. However, it is clear that this desire is also affected by social and educational influences. We are not only born but taught and learn to want to do useful things. Parents teach and influence their children to be so for the good of themselves, their children, and perhaps also of others. Having a desire to achieve is generally good for the welfare of ourselves and of others. However, good qualities may also be detrimental occasionally. Wise people try to avoid this but not all people are perfectly wise. An extreme example of the lack of wisdom is the saying “If I cannot be famous for my

good deeds, I will still try to be notorious for my bad ones". Once a person has a strong desire to achieve, the satisfaction of the desire will make him happy. However, he may be willing to undertake so much hardships that far exceed the happy feeling of achievement. This may still not be irrational if his accomplishments greatly increase the welfare of others. However, most of us have on some occasions been trying to fulfill our desire for accomplishment (or for revenge) in such a way and to an extent that decreases our own welfare without increasing that of others. If this is not just due to ignorance, it is irrational according to my definition. Though this definition may be debatable, we may achieve some agreement by noting that there are different degrees of irrationality. In any case, I do not find that there is any intrinsic normative significance in having the desire for accomplishment fulfilled, except for the welfare effects of such fulfillment, including the indirect ones through the accomplished deeds. Other autonomous desires (such as preferring a less happy "real-world" life to one of being attached to a pleasure machine) may be similarly analysed.

Now, consider a concrete real example. As reported in *Ming Pao Daily* (a reliable leading daily in Hong Kong) on 23.3.97 (p. A11), a man in Tienjin was sent to a hospital after fainting while cycling. Further investigations revealed that he decided to buy a hand phone costing \$9,000+ despite his monthly salary of only \$600+ and his life saving of only \$5,000 (all in Chinese dollars). He thus cut down on all his expenses including food. After more than four months of semi-fasting, he managed to buy the hand phone with a loan of another \$2,000 from relatives. The phone was not for any business or other essential use. Rather, he used it to show off to his friends, cycling from one house to another, ending up in the hospital. Maybe he was ignorant of the possibility of fainting. However, even if he did not faint, I do not think that the happiness he would obtain from showing off his phone would be more than his welfare loss from spending the \$9,000+, including making himself rather unhealthy from semi-fasting. His desire to have the phone, whether autonomous or not, is likely to be irrational.

#### **4. Why is happiness fundamental?**

Why do I regard the satisfaction of preferences, desires, etc. as such not of normative significance while happiness is? Why is happiness fundamental while other things important only to the extent that they directly or indirectly contribute to happiness? The simple answer is that happiness/unhappiness is good/bad in itself and no other thing is so in itself.

There are many things we want: money, job security, status, freedom, etc. However, we do not want them for themselves, but to make us more happy or less unhappy. But we want happiness for itself. It is true that, being happy may also have some instrumental values such as making us healthier and/or more successful in our jobs. However, being healthier or being more successful in jobs are, ultimately speaking, only valuable (in the normative

sense) by contributing to happiness directly or indirectly. Happiness is itself valuable without having to contribute to anything else. If my happiness does not make any other sentient less happy, it is valuable. I do not have to argue with philosophers for thousand of years to establish this since I can directly perceive the enjoyable/painful feelings and know that they are intrinsically good/bad in themselves. But for extreme philosophical solipsists, no one will object to my presumption that other normal individuals have the same capacity for such enjoyment/suffering. (To be less controversial, let us ignore animal welfare here.) Their enjoyment and suffering are intrinsically good/bad to themselves.

To see that happiness is more fundamental than preference, consider advanced computers in the 21st or 22nd century that have preferences but no affective subjective feelings. Clearly their preferences should not count morally. If it is replied that only human (informed) preferences should count, not machine preferences, then consider animals now and advanced computers in the 25th century that do have subjective affective feelings, i.e. they have pain, joy, etc., then most morally sensitive persons will agree that their welfare should also count. Thus, clearly welfare is more important and fundamental than preferences, informed or not, ultimately speaking.

It is uncontroversial that happiness is intrinsically good. The controversial part is saying that it is the only thing that is intrinsically good. I have argued for this elsewhere (Ng 1981, 1990). Recently, I made an additional point to this argument based on the evolution/development of moral principle in a letter to a friend. The content of this appears as Appendix B below. Here, I just want to emphasize the following points. If we follow (as I largely do) Harsanyi in rejecting actual and opting for informed preferences, it is difficult not to go all the way to happiness, in contrast to preferences, as what is ultimately normatively valuable. This is particularly so if we realize that much of our actual and informed preferences and/or our actual behaviour are shaped by our genes to increase our biological fitness which may be at variance with our actual welfare. As pointed out above, it is our welfare that is normatively valuable rather than the dictates of the unfeeling genes formed by random mutation and natural selection. Lastly, as we can all naturally *feel* that enjoyment/suffering is intrinsically good/bad but cannot naturally feel the same for anything else that may be held to be intrinsically good/bad, the burden of argument rests with those who want to replace/supplement happiness by/with something else as ultimately good/bad.

At the risk of repetition, the argument for recognizing happiness as ultimately the valuable thing at the fundamental moral philosophical level does not rule out the importance of insisting on such useful principles as freedom, democracy, law and order, justice, human rights, etc. at the practical, political, or day-to-day level. However, recognizing the real ultimate objective will help us in making decisions on more fundamental issues like the trade-off of the useful principles especially when they are in conflict with each other, the long-term choices of the appropriate institutions and principles to promote.

## Appendix

### *A. A defence of Harsanyi's utilitarianism result*

A compelling utilitarianism result is the well known impartial observer approach by Harsanyi (1953, 1955). Each individual is to indicate “*what social situation he would choose if he did not know what his personal position would be in the new situation chosen (and in any of its alternatives) but rather an equal chance of obtaining any of the social positions existing in this situation,<sup>1</sup> from the highest down to the lowest*” (Harsanyi 1955, p. 316). If the preferences of the individual satisfy a certain reasonable set of rationality axioms, they must define a cardinal SWF equal to the sum of the utilities of all individuals in the society. I find this argument of Harsanyi compelling. I am also very surprised how Rawls (1971), by using a new name (“veil of ignorance”), can obtain a virtually opposite result to Harsanyi – that of maximizing the utility of the worst-off individual only. (For more details of my argument against Rawls, see Ng 1990.)

The equal chance construction can be justified on the principle of “one person one weight”. Take the two person case of you and I. If I dictate the social choice, I may choose mainly with my utility in mind. This is my choice, not the social choice. If I imagine that I have half the chance of being me and half the chance of being you, then I will like the social choice to maximize the unweighted sum of our utilities as that maximize my expected utility. Given uncertain outcomes, the maximization of expected utility is rational. It is commonly believed that Rawls' result is based on risk aversion while Harsanyi's result ignores risk aversion. This is misleading. We are usually risk averse with respect our income as we have diminishing marginal utility of income. But maximizing expected utility instead of maximizing expected income already allows fully for our risk aversion. Since utility (or welfare, where the differences between the two are ignored here) is already our ultimate objective, we cannot have diminishing marginal utility of utility. To be risk averse with respect to utility can be shown to violate some reasonable axioms (Ng 1984). In fact, even if we grant that it is reasonable to be risk averse with respect to utility, we still cannot get the Rawls result unless the degree of risk aversion is infinite, an absurd assumption.

However, Sen (1986, pp. 112–4) questions the significance of Harsanyi's impersonal observer result on the ground that “There is no *independent* concept of individual utilities [other than the von Neumann-Morgenstern values] of which social welfare is shown to be the sum, and as such the result

---

<sup>1</sup> Or, rather, if he had an equal chance of being “put in the place of” any individual member of the society, with regard not only to his objective social (and economic) conditions, but also to his subjective attitudes and tastes. In other words, he ought to judge the utility of another individual's position not in terms of his own attitudes and tastes but rather in terms of the attitudes and tastes of the individual actually holding this position – Harsanyi's note.

asserts a good deal less than classical utilitarianism does” (p. 1123). This is made much clearer and in much stronger terms by Roemer who asserts that “Harsanyi’s Impartial Observer Theorem has nothing to do with utilitarianism” (Roemer 1996, p. 149). Roemer sustains this strongly worded assertion by arguing that the VNM (von Neumann-Morgenstern) utilities and the fully measurable and comparable utilities needed for social choice need have no relation at all. Where the two sets of utility functions are incompatible with each other, it is easy to show that social decisions in accordance to the sum according to one set will in general differ with those in accordance to another set. This criticism of Harsanyi’s result is answered by my result (Ng 1984) that, using axioms no stronger than those for the expected utility hypothesis, with the recognition of finite sensibility (which is just common-sense and well established by psychologists), the utility function derived by the Neumann-Morgenstern method is the *same* as the subjective utility function of classical utilitarianism and neoclassical economists like Edgeworth (1881). Given this result, Harsanyi’s result does give full support to utilitarianism.

Apart from a few other axioms no stronger than those used by the VNM expected utility hypothesis, I was able to establish the equivalence mainly by the following axiom (which is itself implied by the VNM set of axioms assuming infinite sensibility, making explicit preference coincide with intrinsic preference):

Axiom A:  $\forall r, x, y, z : (rIx \ \& \ zPy)$  implies  $[(r, z; 1/2, 1/2) P (x, y; 1/2, 1/2)]$  where  $r, x, y, z$  are alternative outcomes,  $(r, z; 1/2, 1/2)$  is the lottery with 50/50 chance of obtaining  $r$  and  $z$ ,  $P$  is intrinsic preference (the preference if the individual were infinitely sensitive),  $I$  and  $P$  are explicit indifference and explicit preference respectively. The need to distinguish intrinsic and explicit preference/indifference is due to the recognition of finite sensibility. If the optimal amount of sugar in your coffee is 1.8, you may not tell a difference between 1.8 and 1.79 and fail to register an explicit preference but 1.8 may be intrinsically preferred to 1.79. (See Ng 1975 for more details.) This axiom says that the same 50/50 probability mix of obtaining a (explicitly) preferred and an indifferent outcome must be an intrinsically preferred lottery. This is in the spirit of a semiorder that a preference should outweigh an indifference to give at least an intrinsic preference. This is compelling in an atemporal framework where time and future effects have been abstracted away (on which see Ng 1997). Axiom A then implies that a maximal indifference (continuous with a marginal preference, or Edgeworth’s (1881) *minimum sensible*, or just perceptible increment of pleasure) must be represented by a same positive constant. The VNM utility indices satisfying Axiom A then must be the same as the Edgeworthian cardinal utility indices. It is true that the zero point is left undefined in this. However, for utilitarianism, the zero points do not matter, only utility differences count for making social choice (given the set of population; for cases with a variable population, see Ng 1989b).

Even without using the above rigorous axiomatic justification of the equivalence of the VNM utility and the subjective cardinal utility of the classical utilitarians and neoclassical economists, a commonsense argument may also establish this equivalence. We only need to note that the VNM utility indices do not come from thin air just to represent individual choices over lotteries. Rather, each individual *already* has her subjective cardinal utility function to begin with. When I am faced with choices involving risks, I compare the probabilities and the associated *subjective cardinal* utility gains and losses involved before making a choice. Moreover, I choose to maximize, as far as possible, subject to mistakes (which, together with things like regrets, excitement, give rise to various paradoxes and intransitivities; see, e.g. Munier 1988), my expected subjective cardinal utility. Thus, these subjective cardinal utility functions exist *before* the VNM construction is used. The latter is used to discover the pre-existing individual subjective cardinal utility functions by observing their choices involving risks. The degree of the pre-existent diminishing/increasing marginal (subjective and cardinal) utility (of income or some other objective indices) determines the degree of risk aversion/preference, not the other way round.

What the “pure representation” economists do is to say that the VNM utility is purely a representation of individual choices involving risks and has nothing to do with the subjective cardinal utility of the same individual. This would be true if individuals did not consult their subjective utility in making choices involving risks. (But then on what basis do they make rational choices involving risks is rather mythical.) There is a sense in which these pure representation economists are right. First, the well-known axioms of the VNM hypothesis does not ensure that individuals do consult their subjective utilities and try to maximize expected subjective utility in making choices involving risks. However, this may be taken as a commonsense requirement for rational choice or, if one wants to be rigorous, Axiom A above may be assumed to ensure this. This then makes Harsanyi’s result a full utilitarian one. Secondly, Roemer (1996, p. 142) is correct in claiming that the knowledge of all individual VNM preferences or utilities does not give us a meaningful way of making interpersonal comparisons [necessary for any SWF, a utilitarian one in particular]. However, Axiom A and the approach of Edgeworth (1881) and Ng (1975, 1996a) do give us interpersonal comparability. Thus, even if one does not concede that Harsanyi (1953) is sufficient for utilitarianism, one must admit that Harsanyi (1953) plus Ng (1975, 1984, 1996a) are.

Harsanyi (1955) has another (weighted) utilitarianism result. Assuming that both individual and social preferences satisfy the expected utility axioms, then some weak version of the Pareto principle (the Pareto indifference rule requiring social indifference when all individuals are indifferent or some slightly stronger version if non-negative weights are required) is sufficient to yield the result that the social preference is represented by a weighted sum of individual VNM utilities. (See also Vickrey 1945.) Sen again plays down on the significance of this aggregation result, saying that it is primarily a

“representation theorem”, not really utilitarianism. Sen has at least two objections. The first is regarding the choice of individual utility indices as already answered above. The second is that the result is within the single-profile framework with a given set of individual preferences. However, Harsanyi’s result is easily generalized to the multi-profile framework, as must already be implicit in Roberts (1980) who establishes corresponding results in both frameworks, and as explicitly shown in Mongin (1994).

After a long evaluation of the Harsanyi-Sen debate, Weymark concludes that, “If utility only has meaning as representation of [ordinal] preference, then Sen is correct in regarding Harsanyi’s theorems as social utility representation theorems . . . If utility does not simply measure preference, Harsanyi’s Impartial Observer Theorem can be interpreted as an axiomatization of utilitarianism provided (i) well-being is cardinally measurable and fully comparable, (ii) each person’s well-being, including that of the impartial observer, is measured by a von Neumann-Morgenstern utility function, and (iii) the Principle of Welfare Identity is satisfied” (Weymark 1991, p. 315). This last principle is compelling in the context as it just requires that the impartial observer’s ordering of the extended lotteries in which he is individual  $i$  for certain agrees with individual  $i$ ’s ordering of the simple lotteries. Point (ii) is satisfied given my 1984 result cited above in reference to Sen’s objection. Point (i) is compelling to assume for the problem of what form of SWF to take as the existence of any reasonable SWF presumes the existence of interpersonally comparable individual cardinal utilities as argued in Section 1. In fact, Weymark (1991, p. 299) himself argues that the ordinal concept of utility cannot “provide an adequate basis for utilitarianism”, which can have “meaning [only] if utility is cardinally measurable and unit comparable” (p. 303). Thus, for the problem of the appropriate form of SWF, Harsanyi (or anyone else) must be granted interpersonal cardinal utilities. Thus, despite these queries regarding the relevance of Harsanyi’s results to utilitarianism, I believe that Harsanyi emerged completely unscathed.

### *B. Are moral principles ultimately valuable: An evolutionary perspective*

Before the evolution or development of morality and the like, we (perhaps still in the form of apes) had no moral or other principles, no concept of commitments and justice, etc. Self-interest dominated totally, though this does not exclude genetically endowed “altruism” for the maximization of inclusive fitness. As we evolved and more and more relied on our high intelligence and social interaction for survival, the instinct for moral feelings also evolved which helped our survival by enhancing cooperation. This was enhanced by learning the importance of such moral practices as honesty in improving our struggle against nature (including wild animals) and against competing human groups. No one can deny that the initial evolution/development of morality must be purely instrumental (in enhancing either our welfare or our surviving and propagation fitness) as there existed no morality

to begin with. We then learned and taught our children and students to value moral principles, etc., first as a way to increase the degree of adherence to these principles and hence our welfare. Eventually, some, if not most, people came to value these principles in themselves by learning and probably also by instinct. The evolution of such commitment enhancing devices as blushing can be fitness-enhancing; see Frank (1987). Failing to see the ultimate values is a kind of illusion fostered by learning (I dare not say indoctrination) and perhaps genetics. However, I personally have great moral respect for people with such illusions. They most probably make better citizens, friends and colleagues. But illusions they are nevertheless, at least at the ultimate analytical or critical level. While on the whole positive (in maintaining the moral standards), such illusions do have some costs in delaying the rejection of certain out-dated moral principles.

Suppose I suddenly become zombie-like with completely no subjective feelings. Then I do not care at all (even before I become zombie-like) whether you will fulfill your commitment to me or not afterwards, provided no one's (including animals') welfare will be reduced directly or indirectly (including through a marginal reduction in the morality standard). Do you really care? The same applies to freedom, justice, etc. To me, everything that has moral compulsion depends ultimately on welfare. Thus, my answer to Amartya Sen's comment on my position is: While welfare is not equivalent to value, all values are ultimately justified upon welfare.

## References

- Dworkin, R. (1977) *Taking Rights Seriously*. Harvard University Press, Cambridge, MA
- Edgeworth, F.Y. (1881) *Mathematical Psychics*. Kegan Paul, London
- Frank, R. (1987) If *homo economicus* could choose his own utility function, would he want one with a conscience? *Amer. Econ. Rev.* 77: 593–604
- Hammond, P.J. (1987/1995) Altruism. In: J. Eatwell, M. Milgate, P. Newman (eds.) *The New Palgrave: A Dictionary of Economics*. London: Macmillan; reprinted in S. Zamagni (ed.) *The Economics of Altruism*. Elgar, Aldershot, UK
- Hamilton, W.D. (1964) The genetical evolution of social behaviour, I and II. *J. Theoret. Biol.* 7: 1–52
- Harrod, R.F. (1948) *Towards a Dynamic Economics*. Macmillan, London
- Harsanyi, J.C. (1953) Cardinal utility in welfare economics and in the theory of risk-taking. *J. Polit. Econ.* 61: 434–435
- Harsanyi, J.C. (1955) Cardinal welfare, individualistic ethics, and interpersonal comparison of utility. *J. Polit. Econ.* 63: 309–321
- Harsanyi, J.C. (1995) A theory of prudential values and a rule utilitarian theory of morality. *Soc. Choice Welfare* 12: 319–333
- Harsanyi, J.C. (1997) Utilities, preferences, and substantive goods. *Soc. Choice Welfare* 14: 129–145
- Hoffman, M.L. (1981) Is altruism part of human nature? *J. Personality Soc. Psychology* 40: 121–137
- Kemp, M.C., Ng, Y.-K. (1976) On the existence of social welfare functions, social orderings and social decision functions. *Economica* 43: 59–66

- Mongin, P. (1994) Harsanyi's aggregation theorem: multi-profile version and unsettled questions. *Soc. Choice Welfare* 11: 331–354
- Monroe, K.R. (1996) *The Heart of Altruism*. Princeton University Press, Princeton
- Munier, B.R. (ed.) (1988) *Risk, Decision and Rationality*. Reidel, Dordrecht
- Ng, Y.-K. (1975) Bentham or Bergson? Finite sensibility, utility functions, and social welfare functions. *Rev. Econ. Studies* 1: 51–68
- Ng, Y.-K. (1979) *Welfare Economics*. Macmillan, London
- Ng, Y.-K. (1981) Wefarism: a defence against Sen's attack. *Econ. J.* 91: 527–530
- Ng, Y.-K. (1983) Some broader issues of social choice. In: Pattanaik, P., Salles, M. (eds.) *Social Choice and Welfare*, pp. 151–173. North-Holland, Amsterdam
- Ng, Y.-K. (1984) Expected subjective utility: Is the Neumann-Morgenstern utility the same as the Neoclassical's? *Soc. Choice Welfare* 1: 177–186
- Ng, Y.-K. (1989a) Individual irrationality and social welfare. *Soc. Choice Welfare* 6: 87–101
- Ng, Y.-K. (1989b) What should we do about future generations? Impossibility of Parfit's Theory X. *Econ. Philos.* 5: 235–253
- Ng, Y.-K. (1990) Welfarism and utilitarianism: a rehabilitation. *Utilitas* 2(2): 171–193
- Ng, Y.-K. (1992) Do individuals optimize in intertemporal consumption/saving decisions? A liberal method to encourage savings. *J. Econ. Behav. Organ.* 17: 101–114
- Ng, Y.-K. (1995) Towards welfare biology: evolutionary economics of animal consciousness and suffering. *Biol. Philos.* 10: 255–285
- Ng, Y.-K. (1996a) Happiness surveys: some comparability issues and an exploratory survey based on just perceivable increments. *Soc. Indicat. Res.* 38: 1–27
- Ng, Y.-K. (1996b) Complex niches favour rational species. *J. Theoret. Biol.* 179: 303–311
- Ng, Y.-K. (1997) *Utilitarianism: A defence*. Typescript
- Pigou, A.C. (1929) *The Economics of Welfare*. Macmillan, London
- Parks, R.P. (1976) An Impossibility theorem for fixed preferences: a dictatorial Bergson-Samuelson welfare function. *RES* 43: 447–450
- Ramsey, F.P. (1928) A mathematical theory of saving. *Econ. J.* 38: 543–559
- Rawls, J. (1971) *A Theory of Justice*. Clarendon Press, Oxford
- Roberts, K.W.S. (1980) Social choice theory: the single-profile and multi-profile approaches. *Rev. Econ. Studies* 47: 441–450
- Roemer, J.E. (1996) *Theories of Distributive Justice*. Harvard University Press, Cambridge, MA
- Sen, A.K. (1970) *Collective Choice and Social Welfare*. North-Holland, Amsterdam
- Sen, A.K. (1986) Social choice theory. In: Arrow, K.J., Intriligator, M.D. (eds.) *Handbook of Mathematical Economics*, Vol. 3, pp. 1073–1181. North-Holland, Amsterdam
- Vickrey, W. (1945) Measuring marginal utilities by reactions to risk. *Econometrica* 13: 319–333
- Weymark, J.A. (1991) A reconsideration of the Harsanyi-Sen debate on utilitarianism. In: Elster, J., Roemer, J.E. (eds.) *Interpersonal Comparisons of Well-Being*. Cambridge University Press, Cambridge
- Wilson, E.O. (1975) *Sociobiology: The New Synthesis*. Harvard University Press, Cambridge, MA