Contents

1
Introduction

Welfare economics is a very important branch of economic theory. It serves as a foundation for many applied (relatively speaking) branches of economics, such as public finance, cost–benefit analysis and the economics of government policy in many areas, including international trade, industry and welfare (social security and so on). The importance of welfare economics is scarcely in any doubt.

Most people would agree with Pigou (1922, who was the first to address welfare economics as an independent area of study), that 'practical usefulness, not necessarily, of course, immediate and direct, but still practical usefulness of some sort' is what we mainly look for in economic investigation. In other words, bearing fruit is more important than just shedding light. To apply economics beneficially in government policies and to solve social issues we need some guidelines or criteria. Most practical policy problems are not simple enough to allow easy answers. For example if a change will increase the national income but make it more unequally distributed, is it desirable? If a policy will make certain groups of people better off and others worse off, should it be adopted? Should government revenue be raised more by direct or by indirect taxes? Should we go for freer trade even if that will lead to the collapse of some industries? Is globalisation desirable? Should we tax or regulate pollution? To what extent should we conserve our scarce resources? Is economic growth a good thing? The study of welfare economics can help us to answer these questions, but just what is welfare economics?

(Beginners may find the methodological discussion in this chapter rather abstract. Nonetheless they are advised to read it and will come to understand it better after they have read a few more chapters. However if they cannot bear to read it, they can restrict their reading to Section 1.5 and go on to Chapter 2 without much loss of continuity. More advanced readers are advised to read the methodological discussion carefully.)

1.1 What is welfare economics?

Welfare economics is a branch of study that endeavours to formulate propositions that enable us to state that social welfare in one economic situation is greater or lesser than in another. This definition is not much different from the following one by Mishan (1969b, p. 13) 'Theoretical welfare economics is… that branch of study which endeavours to formulate propositions by which we may rank, on the scale of better or worse, alternative economic situations open to society.' In fact if we define social welfare as whatever is good, or whatever ought to be maximised, then the two definitions are identical. However the terms 'better' and 'worse' are explicitly normative, while 'social welfare' may be given a normative or a positive interpretation. It is true that most people tend to regard social welfare as a normative term, but there is no logical reason why we cannot adopt a positive definition of it. Two such definitions are presented below.

First, social welfare can be defined as a vector of individual welfares:

$$W = W(W^1, W^2, \ldots, W^I) \tag{1.1a}$$

where W^i is the welfare of the ith individual and I is the relevant number of individuals. Here individual welfare can be taken as an individual's well-being, or more explicitly, his or her happiness, with happiness subsuming both sensual pleasure and pain and spiritual delight and suffering. But how can individual (net) happiness be measured? One way of escaping this difficulty is to assume that individuals are the best judges of their own welfare and that they maximise this welfare. So whenever they prefer x to y they are assumed to be happier at x than at y. We can then use their utility function (which represents their preference) as an ordinal indicator of their welfare (on ordinal versus cardinal measurability, see Section 1.4). Alternatively we can define social welfare as a vector of individual (ordinal) utilities. One way or another, we have

$$W = (U^1, U^2, \ldots, U^I), \tag{1.1b}$$

where U^i is a utility function representing the ordinal preference of individual i. (We are not concerned here with the technical questions of the conditions that are necessary or sufficient for such a representation; for this see Ng, 1979/83, app. 1B.)

A vector is said to be larger than another if and only if some of its elements are larger than and none of its elements is smaller than the corresponding elements of the other vector. Thus if we define social welfare as a vector of individual welfares (or utilities), we say that social welfare increases if and only if W^i (or U^i) increases for some i and decreases for no i. If welfare increases for some individual and decreases for some other individual, the

change in social welfare (according to the vector definition) is undefined in sign and magnitude.

The vector concept of social welfare must be carefully distinguished from the concept of a Paretian social welfare function (SWF). The Pareto criterion says that social welfare increases if some individuals are made better off without any individual being made worse off, where 'better off' means 'happier' or 'in a more preferred situation'. A Paretian SWF accepts the Pareto criterion. Hence an increase in some W^i (or U^i) and a decrease in no W^i (or U^i) is a sufficient but not a necessary condition for an increase in social welfare. For example, for individuals to live in America it is sufficient that they live in New York, but it is not necessary for them to live in New York – they may instead live in Washington, which is also in America. Similarly if a change satisfies the Pareto criterion it must be regarded as a good change according to a Paretian SWF. But a change need not necessarily satisfy the Pareto criterion to be regarded as a good change. For example a change may make a few individuals marginally worse off but many individuals significantly better off, so it can be regarded as a good change by a Paretian SWF. A Paretian SWF can be written as

$$W = f(W^1, W^2, \ldots, W^I) \tag{1.2a}$$

$$\partial f / \partial W^i > 0 \text{ for all } i \tag{1.2b}$$

Equation 1.2a is an individualistic Bergson SWF (Bergson, 1938) and Equation 1.2b makes it Paretian. By the definition of a function, there exists only one value of W for each set of values of W^i, $i = 1, \ldots, I$. Thus if we have a (specific and fully defined) Paretian SWF we know that social welfare in an alternative situation is greater or lesser even if some W^i vary in opposite directions compared with the original situation. But for the vector concept of social welfare such a comparison is not available.

The vector concept of social welfare is of course of limited interest due to its avoidance of interpersonal comparisons of welfare or utility. Most people accept the Pareto criterion as a sufficient but not necessary condition for an increase in social welfare. But it is difficult to get people to agree on a specific Paretian SWF or to provide the necessary and sufficient condition for an increase in social welfare. Hence what is generally accepted is a vague, unspecified Paretian SWF in the form of Equation 1.2 but with the precise form of f unknown. Hence the vector concept of social welfare in a sense captures the 'minimum content' of this agreement. For example an analysis that deals only with the *necessary* conditions for Pareto optimality may be based on the vector concept of social welfare only. We can then say that the vector social welfare is not maximised unless such and such hold, whereupon the analysis does not have to be based even on the existence of a general unspecified form of SWF, Paretian or not, and one need not be concerned

with the conditions for the existence of a mathematical function. Moreover even people who do not accept the Pareto value judgment can agree that the analysis has some objective meaning. This makes it possible to interpret welfare economics as a positive study.

Another positive definition of social welfare is the utilitarian concept of the sum total of individual happiness:

$$W = W^1 + W^2 + \ldots + W^I = \sum_{i=I}^{I} W^i \tag{1.3a}$$

or if a more objective indicator is desired, one may prefer:

$$W = U^1 + U^2 + \ldots + U^I = \sum_{i=I}^{I} U^i \tag{1.3b}$$

The advantage of adopting Equation 1.3 instead of Equation 1.1 is that with Equation 1.3 social welfare is not incomparable if some W^i increase and some decrease. A difficulty with Equation 1.3 is the problem of interpersonal comparison of welfare or utility (see Section 1.4). Since these individual welfare or utility indices are to be summed, we must be able to find a common unit. In other words the utility functions have to be unit comparable (Sen, 1970b). We shall return to this problem later. At the moment it is sufficient to note that while the problem of interpersonal comparability of utility is a tricky one, it is not insoluble in principle (see Section 5.4.1). It is conceivable that, perhaps several hundred (or a thousand) years from now, neurology may have advanced to the stage where the level of happiness can be accurately correlated to some cerebral reaction that can be measured by a 'eudaimonometer'. Hence the definition of social welfare in Equation 1.3 is an objective definition, although the objects are the subjective feelings of individuals. Deciding whether a particular dish is delicious is subjective, but the fact that a particular individual enjoys that dish is objective. However before we find a perfect 'eudaimonometer' we may disagree widely about the measurement of W^i or even U^i. But if we adopt an objective definition of individual welfare, or happiness or utility, such disagreement is a difference in the subjective judgment of fact, not a difference in basic value judgment (see Appendix 1.1 on the difference between basic value judgment and subjective judgments of fact). The question of whether we ought to pursue or maximise social welfare, as objectively defined in Equation 1.3, is a value question, but the analysis of objectively defined concepts can proceed with or without agreement on such a value question.

It is true that, unless objectively defined concepts are of some interest, analysis of them, while possible, is of little relevance. For example one may define X as the sum of the square root of the number of hairs of individuals divided by the sum of their bank account numbers, and analyse the factors that affect X. However such exercises are of little interest. Hence one would not build a welfare economics based on a definition of social welfare that

appeals to no one. It will be argued in Section 5.4.1 that the concept of social welfare in Equation 1.3 is consistent with a widely acceptable set of value judgments. It is, however, unlikely that any specific concept of social welfare will find universal acceptance as the right objective to maximise. For example even the Pareto value judgment, which seems so mild and reasonable, has its vehement opponents (although it is likely that the opposition is based on a misunderstanding of the Pareto value judgment – see Section 2.1). Conversely the seemingly crazy objective of maximising the welfare of the worst off (implying a zero trade-off between the welfare of the second worst off and that of the worst off) attracts overwhelming attention and has a sizeable group of adherents. Thus what welfare economists can do is to use either a concept of social welfare that they believe to be the right objective, the one that most people or the government believe to be so, or some compromise. This is not very different from other branches of study. For example one may investigate ways to preserve 'mo-xu-you'. This line of research may prove highly useful if mo-xu-you becomes very scarce relative to demand. Even if mo-xu-you remains a free good, its study may not be of much use but it still constitutes a part of our scientific knowledge.

The advantage of adopting an objective definition of social welfare is that it enables us to regard welfare economics as a positive study. However, whether welfare economics is positive or normative is by no means widely agreed upon in the profession.

1.2 Is welfare economics a positive or normative study?

A positive study asks the question: What is? A normative study asks the question: what ought to be? According to Mill (1844, pp. 123–4), 'These two ideas differ from one another as the understanding differs from the will, or as the indicative mood in grammar differs from the imperative. The one deals in facts, the other in precepts. Science is a collection of truths; art, a body of rules, or directions for conduct.' Positive propositions can be verified or falsified, or at least are verifiable or falsifiable in principle. Normative propositions, on the other hand, cannot be true or false; they can only be persuasive or otherwise.

Now, what about welfare economics? While there is no consensus, a majority of economists seem to regard welfare economics as normative (although a minority emphasise the importance of ethics in economics – see for example Hennipman, 1995; Vickers, 1997). This seems a little curious as a majority also regard economics as a science. If economics is a science (which is positive), then welfare economics, as part of economics, should also be a positive study. But is welfare economics perhaps not part of economics? There is an apparent inconsistency here.

The answer to our question depends on our attitude towards the study of welfare economics. If we define social welfare in some positive sense and

confine ourselves to studying the economic factors that affect social welfare, then it is a positive study. On the other hand, if we want to go a step further and do not confine ourselves to saying that a certain measure will increase social welfare (defined in some positive sense), but try to say that a certain measure should be adopted, then we are adopting normative language. We can, however, avoid making value judgments by instead saying, 'If the objective is such and such, then the measure should be undertaken', without committing ourselves to the value judgments behind the objective function.

One possible objection to our attempt to define the concept of social welfare in a positive sense should be considered. The term social welfare has been so widely used in the normative sense that any attempt to define it positively is more likely to cause confusion than to clarify the issue. To forestall such an objection we can refer to the vector concept of social welfare as the welfare vector and to the Benthamite (that is, utilitarian) concept of social welfare as the welfare aggregate.

Little objection can be raised against the concept of welfare vector, apart, perhaps, from its limited usefulness. The concept of welfare aggregate presents more difficulties. Before we can aggregate individual welfares we must be able to measure them and compare them. This requires us to consider the distinction between welfare and utility before considering the problem of utility measurability and interpersonal comparability.

1.3 Welfare versus utility*

It was stated above that one way to reduce the difficulty of measuring individuals' welfare is to take their preference as an indicator of their welfare, such that whenever they prefer x to y we infer that their welfare is higher in x than in y. There are, however, three reasons why this may not always be a good indicator.

First, preference may differ from welfare due to ignorance and imperfect foresight. While individuals may prefer x to y believing that they will be better off in x than in y, it may turn out to be the other way round. This is the question of *ex ante* estimate versus *ex post* welfare. (The recent distinction by Kahneman *et al.*, 1997, between decision utility and experienced utility is based on the same difference.) While the *ex ante* concept is relevant for explaining behaviour, it is the *ex post* one that is actual welfare. Thus Harsanyi (1997) emphasises that informed preferences should be used for normative purposes instead of actual preferences.

Second, the preference of individuals may be affected not only by their own welfare but also by their consideration for the welfare of others. Thus it is possible for individuals to prefer x to y and yet be less happy in x than in y because they believe that other people are happier in x than in y. While it is true that the belief that other people are happy may make them happy, this may not be strong enough to outweigh the loss they suffer from

changing from *y* to *x*. For example individuals may vote for party *x*, even though they know that they will be better off with party *y* in government. The reason why they vote for *x* is that they believe that the majority of the people will be much better off with *x*. This may make them feel better (affective altruism) and is a type of external effect. Although this external benefit may not be important enough to compensate for, in terms of subjective happiness, their personal loss, say in income, under *x*, they may yet vote for *x* due to their moral concern (non-affective altruism) for the majority. To give an even more dramatic example, consider a group of individuals who expect to lead a very happy life. If their country is invaded they may volunteer for a mission that is certain to result in their death. The prospect of being citizens of a conquered country – especially when added to the guilty conscience that would come from failing to volunteer for the mission – may not be too bright, but overall they might still be fairly happy to lead such a life. Yet they choose death for the sake of their fellow countrymen. They do not maximise their own welfare. (This divergence between welfare and preference due to consideration for others is discussed in Ng, 1969b, p. 43; Ng, 1999a; Sen, 1973c.)

Some economists have difficulty seeing the distinction between preference and welfare, saying that whenever individuals prefer *x* to *y* they must be, or at least believe themselves to be, happier in *x* than in *y*. This difficulty is completely baffling. Clearly parents often sacrifice their happiness for the welfare of their children, and it is not easy to see why similar sacrifices cannot be made for a friend, relative, fellow countryman, any other human being on any sentient creature. (For interviews with real-life altruists, see Monroe, 1996, Part I. For a survey of evidence of true altruism, see Hoffman, 1981. One piece of evidence is that a person is more likely to help someone when he or she is the only other person around, contrary to the egoistic explanation that help is given merely to gain the approval of onlookers. Cf. Charness and Rabin, 2002.)

It may be doubted that the existence of true, non-affective altruism is consistent with Darwinian natural selection. However preferences are the result of both cultural and genetic inheritance, and it has been demonstrated that prosocial traits might well have evolved under the joint influence of cultural and genetic transmission (Boyd and Richerson, 1985; Sober and Wilson, 1998; Bowles, 2000).[1]

1. Moreover, 'highly developed human capacities for insider-outsider distinctions and cultural uniformity within communities greatly increase the likely importance of group selection of genetically transmitted traits and hence the evolutionary viability of group-beneficial traits' (Bowles and Gintis, 2000, p. 1419). On the evolutionary basis of altruism towards one's relatives, see Hamilton (1964) and Bergstrom (1996).

If some readers doubt the existence of truly non-affective altruism, they may be convinced that in fact they themselves possess some degree of non-affective altruism if they consider the following hypothetical choice. Suppose that you are asked by the Devil to press either button A or button B within two seconds. You know with certainty that one of the following will happen depending on which button you press. During the two seconds you will be so preoccupied with pressing the right button that your welfare will be zero whichever button you press. After pressing you will lose your memory of the present world and hence will feel no guilt, warm glow or the like, irrespective of which button you press.

- Button A: you will go to Bliss with a welfare level of 1 000 000 trillion units. Everyone else will go to Hell with a welfare level of minus 1 000 000 trillion units each.
- Button B: you will go to Bliss Minus with a welfare level of 999 999 trillion units. Everyone else will go Niceland with a welfare level of 999 trillion units each.
- If you press neither button within the two seconds, you and everyone else will go to Hell.

By construction, choosing A will maximize your welfare but most people will choose B out of non-affective altruism. If you still think that you will choose A, change Bliss Minus into a welfare level of 999 999.999999999 trillion units. If you still opt for A, it has to be conceded that you are not non-affectively altruistic. But how could you have the heart to condemn all others to Hell for a fractional increase in your own welfare? (In my view the existence and degree of non-affective altruism marks true morality.)

Assuming that the preference of an individual can be represented by a utility function (on which see Ng, 1979/83, app. 1B), the difference between welfare and preference discussed above may be illustrated thus. The preference of a rational individual with perfect knowledge (irrationality will be discussed below) is in general a function of the welfare of all individuals.[2] It terms of her or his utility function, we have:

---

2. While Equation 1.4 is a much more general function than the usual utility function, it still involves some simplification. Apart from its incomplete account of irrationality, it also largely ignores such factors as agency (one may prefer to be the one who helps to increase the welfare of others; see Sen, 1987), though to some extent such factors could be partially accounted for by interpreting some of the *x*'s in Equation 1.5 as donations, activities that help others and so on (but the causal connections between such activities and the welfare of others are difficult to reflect in the simplified formulation in the text). This simplification is similar to that of the usual utility function, where only the amounts of goods consumed are important, not how they are acquired and so on.

$$U^i = U^i\,(W^1,\, W^2, \ldots, W^I) \tag{1.4}$$

In general it is not true that individual i prefers x to y if and only if $W^i(x) > W^i(y)$. This is so for the special (though it may be a very important) case of a 'self-concerned' individual (who has no non-affective altruism), where $U^i = U^i(W^i)$. Even for such a self-concerned individual, his or her welfare may still assume the following general form

$$\begin{aligned} W^i &= W^i(x_i^1, \ldots, x_G^j, x_i^2, \ldots, x_G^2, \ldots, x_G^I, W^1, \ldots, W^{i-1}, W^{i+1}, \ldots, W^I) \\ &\equiv W^i(x_{g=1,\ldots,G}^{i=1,\ldots,I};\, W^{j\neq i}), \end{aligned} \tag{1.5}$$

where x_g^j is the value of the gth variable (good, service or activity) by the jth individual, and $x_{g=1,\ldots,G}^{i=1,\ldots,I}$ is just a shorthand way of writing the I times G variables. It should be noted that $\partial w^i/\partial x_g^i$ in Equation 1.5 includes only the direct effect – for example the effect of your smoking on my health – and does not include the indirect effect through W^j, since that is included in $(\partial w^i/\partial w^j)(\partial w^j/\partial x_g^j)$, for example the effect of your smoking on your health (hence welfare), which affects my welfare if I mind about your welfare.

Self-concerned individuals must be distinguished from 'self-minded' and 'self-attending' ones. Self-concerned people have no concern for the welfare of others except insofar as their own welfare is affected. The welfare (but not necessarily also the preference) of self-minded individuals is not affected by the *welfare* of others. The welfare of self-attending individuals is not affected by the *activities* of others. Individuals who are both self-minded and self-attending can be called 'self-regarding'.

In our world of pervasive interdependence the existence of a truly self-attending person is doubtful, though for certain problems it can be assumed for analytical simplicity without much loss. Non-self-concerned (if to do with positive concern) can also be termed non-affective altruism, while non-self-minded can be called affective altruism. Self-concerned individuals are more likely to exist (it could be suggested that over 90 per cent of people are 99.9 percent self-concerned except with respect to their immediate family.) Generous people who help others a lot are not necessarily non-self-concerned since they may only be providing that help because it makes them feel happy. In other words, affective altruists are more common than non-affective altruists.

A self-centred individual is one who is both self-concerned and self-minding:

$$U^i = U^i\left(W^i\left[X_{g=1,\ldots,G}^{j=1,\ldots,I}\right]\right) = f^i\left(X_{g=1,\ldots,G}^{j=1,\ldots,I}\right). \tag{1.6}$$

If this individual is also self-regarding, he or she can be called extremely self-centred:

$$U^i = U^i\left(W^i\left[X^i_{g=1,\ldots,G}\right]\right) = f^i\left(X^i_{g=1,\ldots,G}\right). \tag{1.7}$$

It might be thought that our classification of self-concerned and so on is rather artificial and useless. However for certain problems the distinction provides insights that have policy implications. For example the presence of non-affective altruism may render the Pareto principle with respect to preference unacceptable (Ng, 2000a, app. D) and either affective and/or non-affective altruism may render the Coase Theorem invalid (see Section 7.4 below).

Third, individuals may have irrational preferences. The preference of individuals is defined as irrational if they prefer x over y despite the fact that their welfare is higher in y than in x, and their preference is unaffected by considerations of the welfare of other individuals (any sentient creature can be an individual here), or by ignorance or imperfect foresight. The definition of irrationality is such as to make the three factors discussed here exhaustive causes of divergence between preference and welfare.

While few if any individuals are perfectly ignorant and irrational, some degrees of ignorance (or imperfect information) and imperfect rationality clearly apply to most individuals (for reviews of the relevant literature in philosophy and psychology see Cohen, 1983; Evans and Over, 1996; Kahneman and Tversky, 1996; Stein, 1996), though some alleged irrationalities could be simply due to errors, computational limitations and incorrect norms by the experimenters (Stanovich and West, 2000). There are a number of causes other than ignorance and a concern for the welfare of others, that may make preferences differ from happiness and hence be irrational according to the definition here. The following two (perhaps not completely independent) causes can both be explained, or at least partly, by biological factors (on the biological basis of social behaviour see Wilson, 1975; Crawford and Kreps, 1998).

First, there is the tendency among many people to discount the future or even to ignore it completely. This tendency has been 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, and Harrod (1948, p. 40) regarded it as the 'conquest of reason by passion'. Discounting future consumption, income and any other monetary value is rational as a dollar now can be transformed into more than a dollar in the future. Discounting future utility may still be rational if realisation 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 such irrationality is insufficient savings for old age, necessitating compulsory

and heavily subsidised superannuation schemes. The present author came across an extreme example of the failure to provide properly for retirement during a survey on how much more people would be willing to save if the rate of interest were higher (Ng, 1992a). The question implicitly assumed that everyone did some saving, as the answers were in terms of the extra percentage that would be saved. One subject declared that he did not save anything. He was then asked to change the answer 'saving 20 percent more' into 'saving $20 more per month' but he still said he could not be persuaded to save anything no matter what the interest rate (500 per cent was mentioned). He only conceded a willingness to save when he was asked 'What if a dollar saved now becomes a million dollars next year?' I was careful to find out that this healthy looking young man was not expecting an early death from a terminal disease or the like.

The behaviour of most other animals is largely determined by preprogrammed instincts rather than a careful calculation of present costs versus future benefits. The storing of food by ants, squirrels and so on is 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, such as fight or flight. The ability to anticipate rewards in the fairly distant future requires much more reason, imagination and 'telescopic faculty'. While we know that we are endowed with such a faculty, because it is virtually absent in most other species it is natural to conclude that it is not yet fully developed in our own, and that different members of our species will possess it to differing degrees. Moreover, as Ben-Ner and Putterman (2000, p. 95) put it, 'Given the cumulative character of biological change – that is, its tendency to build opportunistically on preexisting structures – the evolutionary approach leads us to expect that the mind is most unlikely to be a perfect reasoning machine.'

Second, there is the excessive temptation of pleasure (especially present pleasure versus future costs, which relates this cause to the preceding one) and powerful biological drives. After the evolution of flexible species (defined as those whose behaviour was governed not only by automatic programmed responses but also by choice), natural selection ensured that choices would be consistent with fitness by endowing flexible species with a reward–penalty system. Thus eating when hungry and mating with fertile members of the opposite sex were rewarded with pleasure, and damage to the body was penalised with pain. (This meant that flexible species were also 'rational', as defined by Ng, 1996b, who shows that complex niches favoured rational species that made the environment even more complex, leading to a virtuous cycle that accelerated the rate of evolution. This partly explains the dramatic speed of evolution based mainly on random mutation and natural selection, a speed doubted by creationists.)

On top of *ex post* rewards and penalties, we are endowed with inner drives to satisfy fitness-enhancing functions such as mating. On the whole these

powerful temptations and drives work in the right direction, making us do things that enhance both our biological fitness and own psychological welfare. However, since evolution is largely to do with fitness maximisation and the welfare-enhancing aspect only indirectly enhances fitness, some divergence between our behaviour and our welfare is inevitable, as our behaviour is determined not only by rational calculation but also by programmed inclinations, including drives. (See Ng, 1995, on the divergence between fitness and welfare maximisation, especially with respect to the number of offspring.) It has also been shown that 'wanting' (or preference) and 'liking' (or welfare) are mediated by different neural systems in the brain and are psychologically dissociable from each other. In other words individuals may prefer something without liking it or prefer something more strongly than can be justified by their liking for it, and *vice versa*. In particular, neural sensitisation of the brain's dopamine system by addictive drugs can create an intense want that goes far beyond that which can be explained by a mere liking of the drugs and the need to relieve withdrawal symptoms (see Berridge, 1999, for a review). As an example of powerful drives, adolescent girls and boys are often propelled by their sexual drive to engage in careless sexual acts, despite the risks to their long-term welfare, such as an unwanted pregnancy or the contracting of Aids. While this is partly due to ignorance, the power of biological drives cannot be denied.

Consider a specific example. Suppose that a man agrees that, in the case of choices that involve risks, the correct thing to do is to maximise expected welfare (assuming there are no effects on the welfare of others) and actually does so for most choices. However for choices to do with sex he chooses x over y even though his expected welfare is lower with x than with y and he knows this to be the case. Here x may involve having sex with many persons without clear knowledge of whether they have Aids (this knowledge, it is assumed, is not feasible to obtain and hence not relevant). His (expected) welfare-reducing choice of x may be due to the biological inclination to have many sexual encounters. He knows that doing so gives him a not 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 which he knows to be of lower expected welfare. (This is not really a hypothetical example. It can be confidently stated that, out of 100 average adult males, at least 10 have actually made such a choice. If one wants more solid evidence, one only has to look at the frequency of prostitution and extramarital sex.) Should we call this preference informed as the person has all the relevant feasible information, or not informed because it is not in accord with his real interests?

The above two causes of irrational preference illustrate the point that, due either to imperfections in our endowed faculties or to the biological bias in favour of reproductive fitness, we may do things that not are consistent with our welfare. The question here is, for normative purposes should we

use welfare or actual preferences/behaviour? Clearly we should use welfare instead of behaviour dictated by biological fitness. According to an old Chinese dictum, 'Out of the three unfilial acts, not having offspring is the greatest'. However for the human species as a whole the population is certainly not getting smaller, and any long-term social welfare function that accounts for the welfare of future generations should account for that. If we were to go for biological fitness, we would prefer unlimited procreation to a smaller population with a higher aggregate welfare even if we would all be suffering with the former option. 'We' are the feeling selves who ultimately care about our welfare (positive minus negative affective feelings). 'We' are not 'them', the unfeeling genes that urge us to maximise our reproductive fitness. Unlike species that are almost completely controlled by their genes and the environment, we are able to change our reproductive fate by using such measures as birth control. For normative issues it is our welfare, rather than the dictates of unfeeling genes, that should count. (For a survey of different concepts of individual welfare see Ackerman *et al*., 1997.)

In addition to the above biological causes there is another source of imperfect rationality. Individuals may stick rigidly to some habit, custom, principle or the like even if they know that this is detrimental to their present or future welfare and the welfare of others, taking account of all effects and repercussions. Customs, rules, moral principles and so on have a rational basis as they provide simple guides to behaviour that is, or at least on the whole, conducive to social welfare. It would be too cumbersome and time-consuming for individuals to weigh the potential gains and losses in terms of social welfare or their own welfare each time they had to make a decision. Hence they tend to stick to their routines, rules, principles and so on without thinking about the gains and losses. If this occasionally results in decisions that are inconsistent with their welfare and the welfare of others, it can be regarded as a cost of pursuing generally good rules. If, say, there is a change in circumstances, strict adherence to some rules may result in persistent net losses in welfare, taking everything into account. Individuals may stick to these rules without knowing that they are no longer conducive to welfare. Then the divergence between preference and welfare can be attributed to ignorance. If they know this and still stick to the rules, they are irrational.

Many readers may disagree with the definition of irrationality adopted here. For example, suppose a man sticks rigidly to the principle of honesty and would not tell a lie even if that would save his life and contribute to the welfare of others, taking everything else into account. According to our definition here, he is acting irrationally, but for those who see honesty as an ultimate good in itself, he may not be irrational. But are there not occasions when it is better to tell a lie? Should not one lie to an invading army that is known to be cruel and unscrupulous? If we were to press hard enough with such questions, it is likely that most people would ultimately rely on welfare

as the justification for any moral principles such as honesty. Personally, I take the (weighted or unweighted) aggregate welfare of all sentient creatures or a part thereof as the only rational ultimate end (my basic value judgment – see Appendix 1.1), and hence define irrationality accordingly. I am aware of the controversial nature of this definition, but it is not necessary to agree with the definition of irrationality given here to agree with the arguments of this book. If preferred, the word 'irrational', as used here, could be taken as 'irrational according to the objective of welfare maximisation'.

However are moral principles really fundamental? Before the development of morality and the like, we (perhaps still in the form of apes) had no moral or other principles, no concept of commitment and justice, and so on. Self-interest dominated, although this does not preclude the existence of genetically endowed altruism for the maximisation of inclusive fitness. As we evolved and relied more and more on our higher intelligence and social interaction for survival, moral feelings also evolved and helped our survival by encouraging cooperation. This was enhanced by recognising the importance of such moral practices as honesty in aiding our struggle against nature (including wild animals) and competing human groups. No one can deny that the initial evolution/development of morality must have been purely instrumental in enhancing our welfare and/or our survival and reproductive fitness as morality had not previously existed. We then learned and taught our children to value moral principles and so on in order to increase the degree of adherence to these principles and hence improve our welfare. Eventually some if not most people came to value these principles for themselves, by learning and probably also instinctively. The evolution of such commitment-enhancing devices as blushing might have been fitness-enhancing (see Frank, 1987). Failure to see the ultimate values is a kind of illusion fostered by learning (one dare not say indoctrination) and perhaps genetics. However it is possible to have considerable respect for people with such illusions – they probably make better citizens, friends and colleagues. But illusions they are nevertheless, or at least at the analytical or critical level. While on the whole they are positive in that they help maintain moral standards, they do have a cost in that they delay the rejection of certain outdated moral principles.

One real-world example where violation of the preferences of people actually improved their welfare happened in the mid 1960s in Singapore under Lee Kuan Yew's government. Lee decided to expropriate a cemetery and use the land for public development. The human remains were exhumed and reburied elsewhere. This could be regarded as extreme violation of the sanctity of the dead and most descendants would not accept a small fortune in compensation for it. Even if the government had only to pay a small fraction of the willingness to accept, the sum would certainly have turned out to be prohibitive. However it can be argued that Lee was

right to attend to the welfare of the living rather than the dead, even if this went against people's wishes, failed to pass the traditional cost–benefit test based on preferences, and failed to to pass a democratic vote (also based on preferences).

It is interesting to examine why people's preference should not prevail in this case. First, it was partly due to the external costs created by excessive respect for the sanctity of the dead. An individual who failed to show due respect would have run the risk of social censure. Due respect for the dead may have served a useful function but it had become excessive due to complex social customs, including the individually rational but socially harmful strategy of pretending to be very respectful. (If this reason can be explained by traditional analysis in terms of external costs, the next one cannot.) Second, even abstracting away the danger of social disrespect, individuals might have genuinely found it unacceptable for the remains of their relatives to be disturbed. However as the order for compulsory exhumation was made by the government they would have accepted it as unavoidable and beyond their control, and hence would have suffered little loss in welfare. It was thus more than a publicness problem. If the decision had been put to the vote, most of them would have felt compelled by respect for the dead to vote against development of the site. But as the decision had been made for them by the government their distress might not have been as great. Thus Lee's decision almost certainly increased social welfare despite being against the preferences of the people. (However this very exceptional example does not justify autocratic decisions against the will of people in general.)

An interesting question arises as to how we should classify the second aspect of the above example according to our tripartite classification (imperfect knowledge, concern for the welfare of others, and imperfect rationality). It could be thought that, provided that the dead are included under 'others', it should be classified as concern for the welfare of others. However until there is evidence to convince us otherwise, it can be said that the dead are not capable of having welfare. Hence it should be classified as imperfect rationality. According to our definition of rationality it is not (perfectly) rational to put respect for the dead over and above the welfare of the living, apart from such concerns as fear of social ostracisation.

Despite the above discussion of divergence between welfare and utility, it is convenient to ignore such divergences except when we come to discuss problems where divergence is important (such as in the case of merit goods and materialistic bias, as discussed in Section 12.3 and Chapter 11 respectively). In other words, in the absence of specific evidence/considerations to the contrary we shall assume that, as a rule, all individuals are the best judges of their own welfare and choose to maximise their welfare. Then the question of welfare measurability coincides with that of utility measurability, to which we now turn.

1.4 Utility measurability and interpersonal comparability

Many students of economics have at some stage been baffled by the controversy over whether utility is measurable or not measurable, cardinally measurable or just ordinally measurable. Ordinal measurability involves the ability to rank. One can say that utility at x is higher than that at y, but one cannot specify how much higher, or compare the differences in utility. Thus one cannot say whether utility at x is higher than that at y by an amount more or less than the amount the utility at y is higher than that at z. If utility is just ordinally measurable, the utility function is said to be unique up to a positive monotonic transformation, since any positive monotonic transformation of a function ($f = g(U)$, $g' > 0$) leaves the ranking unchanged. On the other hand the measurability of utility *differences* makes the utility function unique up to a positive affine transformation, sometimes called a linear transformation ($f = a + bU$, where a and b are constants and b *is* positive). This transformation leaves the proportions of utility differences unchanged. With full cardinal measurability the only permissible transformation is a positive proportionate one (*Let* $f = bU$, where b *is* a positive constant). One can then say how many times utility at x is equal to that at y and also know what corresponds to zero utility.

The confusion over utility measurability is partly due to the use of the term 'utility' both as a measure of subjective satisfaction and as an indicator of objective choice or preference. Another source of confusion is the insufficient distinction between measurability in principle and measurability in practice. In the case of utility as a measure of the subjective satisfaction of an individual, it seems clear that it is cardinally measurable in principle, though in practice it can be very difficult. The difficulties include inaccuracies and possible dishonesty in preference revelation. Moreover even the individual concerned may find it difficult to give a precise measure. For example I prefer grapefruit to oranges and prefer oranges to apples. If you ask me, 'Do you prefer a grapefruit to an orange more strongly than an orange to an apple? I will say, 'That depends on what kinds of fruit I have eaten recently, and on what sort of meal I am having'. If all these are known, then I will be able to give a definite answer. Thus, subject to practical difficulties, my subjective utility is cardinally measurable. If it was just ordinally measurable I would not just have difficulty answering the question, I would also dismiss it as meaningless. It seems clear that any individual will be able to compare the difference in subjective utility between having an apple and an orange and that between having an orange and a house, and be able to compare the difference in subjective disutility between the bite of an ant and the sting of a bee and that between a sting of a bee and having the right arm cut off.

It also seems meaningful to say that I was at least twice as happy in 2002 as in 1992. If I have a perfect memory, I may even be able to pin down

the ratio of happiness to, say, 2.8. It also seems sensible for someone to say, 'Had I known the suffering I would undergo I would have committed suicide long ago', or 'If I had to lead such a miserable life, I would wish I had never been born'. Hence, it makes sense to speak of negative or positive utility. Somewhere in the middle there is something corresponding to zero utility. According to Armstrong (1951, p. 269),

> There can be little doubt that an individual, apart from his attitude of preference or indifference to a pair of alternatives, may also desire an alternative not in the sense of preferring it to some other alternative, or may have an aversion towards it not in the sense of contra-preferring it to some other alternative. There seem to be pleasant situations that are intrinsically desirable and painful situations that are intrinsically repugnant. It does not seem unreasonable to postulate that welfare is $+ve$ in the former case and $-ve$ in the latter.

Hence it seems clear that utility or welfare as a subjective feeling is in principle measurable in a full cardinal sense.

On the other hand we can use a utility function purely as an objective indicator of individuals' preference ordering, such that $U(x) > U(y)$ if and only if they prefer x to y and $U(x) = U(y)$ if and only if they are indifferent; and we may not be interested in anything other than the above ordinal aspect of the utility function. Then any monotonically increasing transformation of a valid utility function is an acceptable indicator and a utility function possesses only ordinal significance. For some problems (such as the theory of consumer choice), knowledge of the preference orderings is all that is required and hence an analyst can abstract away the cardinal aspect of the utility function. This, however, does not mean that for problems where the intensity of preference is relevant (such as social choice), one cannot adopt cardinal utility functions, provided due attention is paid to the practical difficulties. To deny the use of cardinal utility is to commit what could be called the 'fallacy of misplaced abstraction'. The technical problems with the existence of a utility function that represents an individual's preference are discussed in Ng (1979/83, app. 1B). Here we shall, for the most part, take the existence of a utility function for granted. We now turn to the problem of interpersonal comparisons of utility.

Different types and degrees of comparability can be distinguished. Level comparability refers to the comparison of total utilities; it answers questions such as 'Is person J happier than person K?' Unit comparability refers to differences in utilities between situations. It answers questions such as 'Is J made better off by more than K is made worse off by moving from x to y?' Full comparability subsumes the possibility of both level comparability and unit comparability. Non-comparability excludes both. In between lies partial comparability (Sen, 1970a, p. 99, 1970b) where rough, imprecise

comparisons of units or levels of utilities can be made. Different social welfare functions and different problems need different types of comparability (Sen, 1974, 1977b). For example if we are interested in maximising the sum of individual utilities given the number of individuals, then what we need is unit comparability. On the other hand, if we adopt the Rawlsian criterion of maximising the welfare of the worst-off individual, what we need is level comparability (contrary to common belief, the general possibility of level comparability implies unit comparability under fairly general conditions – see Ng, 1984c).

There is a long tradition among economists to regard statements that involve interpersonal comparisons of utility as value judgments (Robbins, 1932, 1938; Wicksteed, 1933). I have argued elsewhere (Ng, 1972b, 1997; see also Appendix 1.1) that such statements are just subjective judgments of fact. The judgment that J is happier than K and that J's gain in happiness will exceed K's loss of happiness if there is a change from x to y does not imply what ought to be done until an objective function (which necessarily involves value judgments) is specified. Thus if our objective is to maximise the sum of utilities, we choose y; if we want to maximise the utility of the worst-off individual, we choose x (assuming that K is and will remain the worst-off individual). Judgments that involve interpersonal comparisons of utility are subjective judgments of fact even though the facts are the subjective feelings of individuals. However, due to this subjective nature it is very difficult to measure individual utilities and to compare them interpersonally with any degree of precision. While such difficulties should not be underestimated, they do not make interpersonal comparisons value judgments.

1.5 The organisation of the book

The rest of this book is organised as follows. Chapter 2 is devoted to Pareto optimality and emphasises the *necessary* conditions for optimality. Chapter 3, on welfare criteria, is concerned with the *sufficient* conditions for social improvement (qualitative), and in Chapter 4 the more quantitative topic of consumer surplus is discussed. Chapters 5 and 6 address the issues of social choice and income distribution, respectively. In Chapter 7 the focus shifts to market imperfections. Perhaps the most important of these is externality (including pollution – students of mine who smoked stopped doing so after this topic had been covered). A special category of externality – public goods – is considered in Chapter 8. When discussing market imperfections the question of second best arises when some of the optimal conditions cannot be met. The theory of second best, as well as the present author's theory of third best and its application to derive a principle of treating a dollar as a dollar in specific issues, are discussed in Chapter 9. Chapter 10 discusses welfare economic issues raised by the Yang–Ng framework when

analysing the division of labour and the evolution of economic organisations. Chapter 11 attempts to push welfare analysis from the level of preferences to that of happiness. Finally, Chapter 12 touches on wider welfare issues and speculates about the need for an interdisciplinary study of happiness.

As outlined in the Preface, some of the chapter sections and appendices are marked with asterisks. In the main, the ones without asterisks contain basic material that is essential for beginners. The ones with a single asterisk are a little more advanced and are likely to appeal to those who already have a basic knowledge of welfare economics. Those with two asterisks contain new arguments of interest to the expert (although beginners may attempt to read all the sections and appendices if they do not find them too demanding). The attention of experts is also drawn to the following: the methodological issues discussed in this chapter and its appendix; in Chapter 3, the assessment of Little and Mishan's argument with respect to welfare criteria (Sections 3.2.2 and 3.3), and the quasi-Pareto criterion proposed by the present author (Section 3.5); in Chapter 4, Appendix 4.2; in Chapter 8, overestimation of the costs of public spending (Section 8.2), and the Clarke–Groves incentive-compatible mechanism for preference revelation (Section 8.3); in Chapter 9, the theory of third best (Section 9.4); and the entirety of Chapters 10 and 11.

1.6 Summary

Welfare economics is a branch of study that endeavours to formulate propositions that enable us to state that social welfare in one economic situation is greater or lesser than in another. It can be regarded as a positive study if a positive definition of social welfare is adapted or if the social welfare function is taken as given. Usually, social welfare is taken as some function of individual welfares or utilities. Welfare may diverge from utility (or preference) due to a concern for the welfare of others, or to ignorance or irrationality. Subject to practical difficulties, individual utility can be cardinally measured; interpersonal comparisons of utility are not value judgments. Distinguishing basic value judgments from subjective judgments of fact enhances the role of economists in policy recommendation (Appendix 1.1).

Appendix 1.1: Basic Value Judgments and Subjective Judgments of Fact

When there is a disagreement about policy matters, about what ought to be done, it is usually said that this is a matter of different value judgments on which no objective discussion is possible. In the second section it will be shown that this is not usually true, but first we shall consider a revised (or sister) definition of Sen's concept of basic value judgment.

A1.1.1 Positive statements versus value judgments

Positive statements are concerned with what is, what was or what will be in the objective sense; normative statements (or value judgments) are concerned with what ought to be, what is morally right or wrong, good or bad. This is a sufficiently clear distinction for our purposes. It is true that the philosophical problems of metanormative theory are by no means completely settled; there is much scope for disagreement about whether the statement 'this picture is beautiful' is positive or normative. (I regard the statement as positive if it means 'this picture appears beautiful to me', as evaluative if it means 'I think this is a beautiful picture', and as normative if it means 'this picture *is* beautiful!', implying that others should accept it as beautiful.) But for the present purpose the finer distinctions between the evaluative, the ethical and the normative are not of direct importance and will therefore be disregarded. What will be emphasised is the distinction between basic and non-basic (or derived) value judgments.

A1.1.2 Basic versus non-basic value judgments*

Sen (1970a, p. 59) defines basic judgments as follows: 'A value judgment can be called "basic" to a person, if the judgment is supposed to apply under all conceivable circumstances, and it is "non-basic" otherwise.' For example individuals may judge that 'An increase in national income indicates a better situation' (judgment A). This is not a basic value judgment if they agree that an increase in the national income would not signify a better situation if the poor were made much poorer. If they were to revise their judgment and say that, 'For a poor country in which everyone had the same income, it would be a better situation, *ceteris paribus*, if everyone's income increased by the same amount' (judgment B), and they stuck to this under all circumstances, should we regard it as their basic value judgment? If we were to ask 'Why do you regard that as a better situation?' and they replied that it was because everyone would become happier, then it would be reasonable to regard judgment B as non-basic (to those individuals).

In the above example the individuals may believe that 'It is desirable to make every individual happier' (judgment C). They stick to judgment B under all circumstances because the conditions 'a poor country', 'equal income', 'equal increments', '*ceteris paribus*' and so on are sufficient (or are believed to be sufficient) to make every individual happier. Thus judgment B is derived from value judgment C and from certain factual, positive statements or judgments. We shall therefore call a value judgment basic to individuals if it is not derived from some other value judgment and they believe in it for its own ethical appeal. Whether this definition differs from Sen's depends on the interpretation of 'all conceivable circumstances' (ibid.) In the above example, the individuals' belief in judgment B is based on

judgment C. If we were to ask, 'Will you stick to judgment B even if a policy based on it does not make anyone happier?', then presumably they would have to say 'No' (I owe this observation to Sen). Thus if circumstances in which acting in accordance with judgment B does not make anyone happier are regarded as conceivable circumstances despite the fact that the individuals believe that they are sufficient to make everyone happier, then Sen's definition may not be different from ours. In fact with this broad interpretation of 'conceivable circumstances' the two definitions are equivalent.

We shall first show that if a value judgment is non-basic in our sense, it is also non-basic in Sen's sense (with the broad interpretation), and then show the reverse. If judgment X is based on judgment Y and some factual knowledge or judgments, this means that, given Y, X applies in the factual domain α. X may or may not apply outside α. Let us expand α to α', within which X applies and beyond which it does not apply. This is just a conceptual expansion and no actual specification of α' is necessary. The universal domain – that is, the domain of all possible (non-zero probability) circumstances – is donated by Ω. If $\Omega - \alpha'$ is not empty, then obviously X is a non-basic judgment in Sen's sense. If α' is itself the universal domain, then $\Omega - \alpha'$ is empty. If we confine 'all conceivable circumstances' to the universal domain, X applies under all conceivable circumstances and is a basic value judgment in Sen's sense. But if we do not confine 'all conceivable circumstances' to the universal domain, then X does not apply beyond α' and is therefore a non-basic judgment in Sen's sense.

If judgment *Z* is not basic in Sen's sense, then it applies in some domain β but does not apply outside that domain. Suppose that judgment T applies outside domain β. It does not matter whether the individuals concerned can make up their mind in a definite way. Thus judgment T may just mean 'No opinion' or 'If..., then...; If..., then...We can then combine Z and T into a single judgment, V. This combination is always possible (on which more below) since V may just stand for 'If β, Z; if non-β, T'. Then clearly judgment Z is based on judgment V and on some factual statements or judgments that delineate β. It is therefore also non-basic in our sense.

From the above it can be said that if a value judgment is non-basic in Sen's sense it is necessarily non-basic in our sense; if it is non-basic in our sense, whether it is also necessarily non-basic in Sen's sense will depend on the interpretation of 'all conceivable circumstances'. Even if we adopt a broad interpretation such that the two definitions coincide, it may still be useful to look at it in our way. Thus in the example above we may not be able to specify a conceivable circumstance in which the individuals will relinquish judgment B until we reveal their more basic value judgment C by asking them 'Why B?'. Similarly, even if we wish to stick to our definition of basic, it may be useful to test for basicness using Sen's method. It is sometimes difficult to ascertain whether a certain value judgment is based on another just by asking 'Why?'

For example individuals may mistakenly but sincerely believe that their judgment 'No one shall kill a human being' (judgment D) is not based on another and is due purely to its ethical appeal. They may not be able to answer the question 'Why do you believe in judgment D?', or they may reply 'I believe in it because I believe in it!' It may therefore seem that judgment D is their basic value judgment. But if we introduce them to the factual circumstances of euthanasia or a fight against an invading army, they may admit that they no longer accept judgment D. Their initial belief that judgment D was basic was thus mistaken. If it had been truly basic, it could not have been shown to be inapplicable under any circumstances.

A1.1.3 Can individuals have more than one basic value judgment?

Suppose an individual has two value judgments that they regard as basic. Then either these judgments are not in conflict with each other under any circumstances or they are under some (or all) circumstances. If the latter is the case, then it is clear that the two judgments cannot both be basic. If the two never give conflicting results, then we can combine them into a single one. For example the judgments 'No female should kill a human being' and 'No male should kill a human being' are never in conflict and can be combined into one, 'No person should kill a human being. An individual may have many diverse basic value judgments that cannot be combined into a single statement and the only way to combine them is to list them exhaustively, for example. 'One should not murder; one should not…' It may then just be a semantic point whether we regard mutually non-conflicting judgments as numerous value judgments or as parts that form a single value judgment or value system. It is even possible that the individuals themselves may think that it is not possible to list all their basic value judgments exhaustively as new ones may crop up, especially with the specification of a new set of circumstances. However in such cases the numerous value judgments are likely to be non-basic. If rational people press themselves hard enough with the question 'Why do I think this is a good thing?', all their supposedly basic value judgments can probably be reduced to a few or just one principle (their basic value judgment) that gives rise to various expressions under various circumstances. Most, if not all, reasonable people will agree that, for society, our ultimate goal should be the maximisation of an increasing function of and only of individual welfare, that is:

$$W = W(W^1, W^2, \ldots, W^I), \frac{\partial W}{\partial W^i} > 0 \qquad \text{(A1.1)}$$

where W indicates social welfare, W^i the welfare of the ith member, and I the number of members concerned. (If we are not using a static analysis then we have to maximise a welfare function through time, discounted to account for the uncertainty of continuing existence. If we are, as we should

be, also concerned with sentient creatures other than human beings, we have to include their welfare.)

However even if we can agree on a specific form of the function, this could hardly serve as our only guide under all circumstances. This is so because it is difficult to know whether a particular action will increase or reduce welfare due to the complexity of the real world. From experience, however, we learn that actions that are in accordance with certain principles (honesty, respecting other people's freedom and so on) usually contribute to welfare. Eventually these principles tend to be valued for their own sake. This may explain the diversity and numerousness of basic value judgments.

A1.1.4 Subjective judgments of fact

The distinction between basic and non-basic value judgments is of some importance, as noted by Sen (1970a, ch. 5). If two persons disagree over a basic value judgment, the only possible scientific (factual, logical) solution is for one to show to the other that the judgment in question is not really a basic value judgment. If it really is basic, no ground is left for scientific argument. One of them may of course try to persuade the other to change his or her basic value judgment. But this is an exercise in ethical persuasion, not scientific discussion. This does not mean that scientists should not engage in such persuasion. Scientists are human beings as much as any other person and are entitled to engage in ethical persuasion provided they do not confuse scientific discussion with ethical persuasion, or attempt to confuse their audience.

On the other hand, if the disagreement is about factual judgments or non-basic value judgments, scientific discussion alone may be sufficient to settle the dispute. However, since our scientific knowledge is never complete a final agreement may not be reached even by two logical persons with the same basic value judgments. This is due to possible differences in 'subjective judgments of fact' (Ng, 1972b). These are estimates made in the absence of definite scientific knowledge and hence may be influenced by the personal interests, values, experience and so on of the person making them. For example it has been observed that, when considering the same statistics, left-wing economists tend to infer that progressive taxation does not have a disincentive effect, while right-wing economists tend to infer that it does (Klappholz, 1964, p. 103). However disagreement about subjective judgments of fact can, or at least in principle, be subject to further discussion and scientific testing. With the advancement of science such differences may eventually disappear, though they are likely to be replaced by others (on consensus among economists, see Kearl *et al*., 1979).

As a scientist, one has no business to make any basic value judgment, and as a citizen a scientist is no more qualified than any other citizen to make basic value judgments. But this is not so for non-basic value judgments and subjective judgments of fact. If we have any faith in the usefulness of

a branch of study, we must admit that scientists in that branch are more qualified than a layperson to make subjective judgments of fact that are closely related to their field of study. For example, while an economist may be hard put to judge the effects on economic growth of two alternative balance-of-payments policies, a layperson is likely to know nothing at all about the effects. Thus the ability to distinguish subjective judgments of fact and non-basic value judgments from basic value judgments increases the role of scientists in policy recommendation. If we were to lump all three under 'value judgments' and refrain, as scientists, from making such judgments, we would have to leave all but undisputable scientific facts to the politicians.

Another source of disagreement about policy matters is differences of interest, rather than differences in values or judgments of fact. For example the leaders of a strong trade union may press for a large wage rise despite knowing that it will lead to inflation and/or unemployment. On the ethical level they may admit (although not in public) that their action is not good for society as a whole. But on the practical level they may be guided mainly by self-interest.

A1.1.5 Summary

To summarise, disagreements (especially on policy matters) may arise as a result of one or more of the following:

- Conflicts between personal or sectional interests.
- Differences in basic value judgments.
- Differences in the subjective judgment of facts.
- Differences over alleged facts or in logical analysis.

Disagreements of the fourth type can readily be resolved by objective demonstration or logical discussion. Any remaining disagreements on this score can only be ascribed to ignorance or the inability to reason. Disagreements of the first type can be resolved by a political compromise of some sort, or by the exercise or threat of force. Alternatively, especially in the long run, they may be resolved by teaching people to be more social-minded or by making their interests more compatible. Disagreements of the second type can only be resolved by ethical persuasion (however see, Piderit, 1993, for the argument that even value judgments are susceptible to rational justification). It is disagreements due to differences in subjective judgments of fact that offer the most scope for further discussion, analysis and argument. Apart from the conflict of interest, it is also, I believe, the most important source of disagreement.

Since a non-basic value judgment is based on a basic value judgment plus factual information or a subjective judgment of fact, differences in value judgments are due either to differences in basic values or to differences

over subjective judgments of fact, ignoring factual and logical mistakes. The latter type of difference (over subjective judgments of fact) tends to be more prevalent and is usually regarded as a direct disagreement about value judgments on which no further argument is possible, as no distinction is made between basic value judgments and subjective judgments of fact. The above discussion sounds a warning against this confusion.

Subject Index