EQUITY IN HEALTH

Alan Williams and Richard Cookson

Centre for Health Economics
University of York
York YO10 5DD
England
EQUITY IN HEALTH

By Alan Williams and Richard Cookson

(about 19,000 words, including references and footnotes)

I SCENE-SETTING

In this chapter our concern is health, and not health care. We view health care as one of many possible determinants of health, and not necessarily the pre-eminent one. We therefore regard equity in access to health care, or in the distribution of health care resources, as an instrumental matter, flowing from a more fundamental concern with the distribution of health itself. We shall assume for the time being that health can be measured unproblematically, and that the unit in which it is measured is the quality-adjusted life year or QALY. We shall restrict our attention to various ways in which an equitable distribution of health within a population might be defined in an operational manner that enables economists to carry out empirical work evaluating possible ways in which it might be brought about. We will assume that in order to achieve a more equitable distribution of health in a population it will generally be necessary to reduce the overall level of health in that population (in other words, we assume that an efficiency sacrifice will usually be required to achieve equity). We will not be concerned with equity in the financing of health care (or in the financing of any other public policy designed to bring about a more equitable distribution of health). Nor will we be concerned here with the use of financial mechanisms explicitly to promote equity in health, important though these may be as instruments as policy. Our purpose is rather to clarify the equity objective itself. For convenience we will assume that the resources the community wishes to be used for the pursuit of health have been set aside in an equitable manner, and our task is to ensure that there are clear criteria for determining how best to use them in pursuit of an equitable distribution of health within that community. The term "community" is used to encompass all those to whom the social decision-makers are responsible.

In economics the term "equity" is usually taken to refer to fairness in the distribution of a good (in this case "health"), and "fairness" is taken almost unthinkingly to mean reducing inequalities. In philosophical writings equity concerns would more likely be broader than this, and called concerns about "distributive justice". Concerns about distributive justice often become intertwined with concerns about procedural justice, a matter which we shall examine

1 See Chapter 34 by Dolan for a general discussion of measurement and valuation issues. We shall return to some of them later in the specific context of equity weights.
2 If this is not the case, the simultaneous pursuit of efficiency and equity is made that much simpler.
3 This is taken up in Chapter 40 by van Doerslaer and Wagstaff.
4 It is possible that the social decision-makers will need to take a broader view of equity when deciding on how much of the community's resources to set aside for health than they need to take into account when focusing on how best to use those resources for health itself (and only for health).
more closely shortly. Philosophical writings tend to focus on what is fair as between individuals with known characteristics. Economists, on the other hand, tend to focus on what is fair as between groups of individuals distinguished only by some common characteristics, accepting that the other characteristics of each individual will be ignored even though they may differ widely. This calls for the exercise of moral sensitivity about "statistical lives" rather than about the lives of named individuals who we can see and touch and talk to. For many people this notion of “statistical compassion” seems to create both intellectual and psychological difficulties. It is as if personal empathy with one or two individuals is possible, but, paradoxically, if many individuals are involved this capacity to empathise diminishes. This difference between focusing on groups and focusing on individuals also distinguishes economists (and managers) from clinicians and others dealing with people at an individual level. The latter often claim that they are under an ethical duty to do everything possible for the person in front of them no matter what the consequences might be for everybody else. If this assertion is taken at its face value, it would imply that clinicians should ignore their responsibilities for the welfare of their other patients except when that patient is in front of them. It seems most unlikely that any clinician would actually behave in that way, so perhaps the statement should not be taken at its face value, but regarded instead as part of the rhetoric of medical practice, designed to bolster the doctor-patient relationship. But whatever may be the role of such statements, it is clear that in a public policy context, where distributive justice is an explicit objective, it is clearly not ethical for a clinician to ignore the consequences of his or her actions concerning the treatment of one patient for the health of others patients for whom the system is also responsible.

The exposition which follows is written primarily for economists with little or no familiarity with the relevant philosophical literature. In order to make this literature more readily accessible, we have selected the main points made by each of the cited authors, and, in the interests of clarity we have stripped away the many qualifications and elaborations that are contained in the original works. These original sources need to be consulted carefully before claiming a proper appreciation of an author’s position. We have also judged it not to be necessary to present detailed practical examples to illustrate each case (which is a common mode of exposition in medical ethics), but have contented ourselves with indicating the kinds of issues addressed by each philosophical theory, leaving the reader to think through the practical implications for particular cases. What we are attempting here is the brutal task of forcing high-minded philosophical theories about distributive justice into the procrustean bed of welfare economics! That painful process will commence with a brief explanation of how we propose to delineate the economist’s conventional conceptual framework for that purpose.

II PHILOSOPHY AND ECONOMICS - AN ANALYTICAL PREAMBLE

In the last twenty years, social science journals have published upwards of one thousand articles on the subject of equity in health (the exact number depending on how widely or narrowly you define that elusive concept). In contrast, the humanities journals contain almost nothing on that topic (or indeed on health more generally). On the rare occasions when philosophers write anything about health, they tend to do so either in social science journals or in books. The voluminous social science literature on equity in health really started to take off in the 1990s. Although the emergence of the World Bank work on DALY’s (World Bank 1993, Murray and Lopez, 1996) was a significant element in this trend, it was part of a more general increase in attention to equity right across the social sciences. A notable feature was the increasing use of the word "fairness", which failed to get a mention in connection with health

---

5 From the standpoint of procedural justice, a distribution of health within a community would be said to be just if it were the result of processes that were just. It would be the processes that would be the object of attention, and not the substantive outcome. The position taken here is that it is the outcome that is the focus of attention, and that the processes are instrumental.
during the 1980s, but logged around 10 mentions a year during the 1990s. By 1997, the most commonly cited philosophical work in articles published in social science journals about equity in health was Rawls' 1971 book "A Theory of Justice" (of which more anon), which did not in fact include health as one of the primary goods which fairness required to be distributed equally. To understand why economists may have had difficulty incorporating philosophical ideas into their thinking, we must first consider carefully what their characteristic mode of thinking is.

Economists typically approach optimisation problems by listing the options to be considered (the "opportunity set") and then choosing between them by applying some maximand (the "objective function"). The opportunity set may be presented as a production possibility frontier, or as a utility possibility frontier. If it is presented as a production possibility frontier, this is the particular set of options (defined as a package of goods and services) that are not "dominated" in the optimisation process by any other option. On and within the frontier, the complete set of options comprises those satisfying two conditions: (i) they must be technically feasible\(^6\) and (ii) they must be producible with the resources available. The resources available can either be represented by a budget constraint (leaving flexible the actual input combinations), or by a fixed allocation of (unpriced) real resources. If, on the other hand, the opportunity set is presented as a utility possibility frontier, the same conditions apply, but the analysis has already proceeded one step further and distributed the goods and services to individuals in different ways so that the options presented are the various interpersonal distributions of utility that result. To analyse the case of equity in health, the analogy with utility possibilities is more appropriate than with production possibilities, since health, like utility, can only subsist in individuals. We shall be presenting the opportunity set as a health possibility frontier which focuses on the interpersonal distribution of health outcomes and assumes it to be measured in terms that are analogous to a utility measure (QALYs).

We shall use this separation of the health possibility set from the social maximand in order to classify and analyse the implications of different theories of justice for the interpersonal distribution of health.\(^7\) We shall argue that some of them (notably those that stress procedural requirements or the primacy of goods other than health) are best seen as restricting, on ethical grounds, eligibility for inclusion in the health possibility set. Others impose special restrictions on the nature of the maximand. Some do both.

As was briefly explained above, the conventional definition of the opportunity set is that in principle it contains every option that is both technically feasible and producible within the resource constraints. But the resource constraints themselves will typically not simply be those set by nature, but also be the result of some human decision or decisions. As such they are likely to reflect the judgements of managerial, professional, technical and political actors concerning what it is wise or reasonable to devote to the objectives in question. Similarly, some things which are technically feasible may be ruled out from further consideration by one or more of these actors because they are not considered sensible or politically feasible within their own particular sphere of authority and expertise. Consequently, in practical terms the opportunity set presented to economists engaged in the evaluation of health policy options may already have been severely truncated by such additional restrictions, and during the process of problem formulation it is important for such investigators to explore how these particular options (and not

---

\(^6\) Technical feasibility is to be understood broadly as any constraint on what can be achieved with given resources, including incentive constraints on individual behaviour (such as those arising from asymmetry of information), as well as constraints relating to the state of technology, more narrowly understood.

\(^7\) Mishan (1977) uses a similar approach, and generates a diagram that is very similar to one of those that we shall be using later.
Here it will be assumed that however the health possibility set was defined, ethical constraints concerning the requirements for equity in health were not part of the process. We shall regard any ethical constraints emerging from theories of justice as additional factors to be taken into account in defining an opportunity set which had previously been innocent of any such considerations. It will furthermore be assumed that such ethical constraints never enlarge the opportunity set. They may leave it untouched, if all the previously considered options happened to fulfill the ethical constraints anyway. They may delete options which were not on the frontier, and which consequently would not have been considered in any case. They may also delete options on the frontier that would not have been chosen, in which case again they are irrelevant. But in general we shall assume that these ethical constraints do "bite", in that they remove from consideration options that would otherwise have been chosen. Otherwise we would be treating them as \textit{a priori} irrelevant.

Theories which restrict the nature of the maximand fall into two groups. The first group establish side-conditions upon outcomes which have to be fulfilled before any maximisation process comes into play. For instance, the side condition might be that there must be some minimum level of health provided for some specified group and no trade-off is permitted between this objective and any other. Once the minimum has been provided, however, the maximand has unrestricted applicability (unless there is a whole sequence of such side-conditions, each of which has to be satisfied in the prescribed order before we get to the residual set to which the maximand is applied). The second group of restrictions upon the nature of the maximand concern its actual content and the weights to be attached to its various elements. These differ from the previous case in that trade-offs are permitted between all of the various ethical desiderata.

In order to deal systematically with these complexities, it will be useful to refer to the accompanying Table, the upper half of which covers those cases where the opportunity set is unaffected by the particular notion of equity under consideration, and the lower half of which covers those cases where it is (including those which affect both the opportunity set and the maximand, such as the Rawlsian equity notions). In each half of the Table there is a further subdivision according to whether the theory imposes side conditions upon outcomes. Where equity is held to require that more than one such condition has to be satisfied, there are two possibilities: one is that they are to be satisfied in a prescribed order, and the other is that they are all to be satisfied simultaneously. In the former case, we first have to ensure (say) that Group A achieves some minimum level of health, and when that has been done we move on to Group B, and so on. In the other case, where all side-conditions have to be satisfied simultaneously, there is no such priority ranking between groups, and all side-conditions have equal salience. In the remainder of the Table it is the particular characteristics of the optimisation criteria are the focus of interest. Under “Nature of Maximand” the first column covers the case where there is no maximand. The remaining columns focus on the shape of the social welfare contours for health. Essentially these may be linear or non-linear, and, if non-linear, they may be smooth or kinked. And within each of these three subgroups, equal weight may be given to each class of

---

8 It might be argued that, paradoxically, ethical constraints can actually enlarge the health possibility set by facilitating trust and co-operation (e.g. between doctors and patients). For instance, an ethical constraint that removed the fear that a certain category of hospitalised patient might be allowed to die, when they would prefer to go on living, simply in order that organs which they had offered to donate might be "harvested" sooner, might make people more willing to donate organs, with consequent additional benefits to others. However, we shall regard such instances as exceptional, and generally regard procedural rules as restricting rather than enlarging the opportunity set.
These are the distinctions that we have found to be important discriminators between the more popular notions of distributive justice.\(^9\) In the analysis which follows, we shall go through the cells in the Table one by one, starting in the top left-hand corner, and proceeding from left to right, one row at a time. Where a cell is empty, we shall comment on what it might have contained had such a notion of justice been propounded by anybody. But the main purpose is to link each notion of equity to a particular analytical device. This will be done by first outlining the theory of justice from which the notion derives, and then showing how it could be represented in welfare economics. For ease of reference each subsection in the text has a summary heading identifying the particular cell whose contents are there being discussed. Further reading on particular theories of justice is suggested in the relevant sections; for a review of economic thinking about equity in health see Culyer and Wagstaff (1993); for a review of economic thinking about equity in general see Hausman and McPherson (1996); for a review of philosophical thinking about equity in health see Daniels (1985); for a review of philosophical thinking about equity in general see Plant (1991).

**III THEORIES WHICH LEAVE THE OPPORTUNITY SET ETHICALLY UNCONSTRAINED**

Before embarking upon that task, however, something needs to be said about the general properties of the health possibility set. Generally we shall assume this to have the properties normally assumed for an opportunity set in welfare economics, namely that it generates a frontier which is continuous and concave to the origin, and monotonically decreasing from left to right. This means that there is a densely packed set of possibilities for generating health for both A and B, but that on the frontier the health of the one can only be improved by sacrificing some of the health of the other. Later we shall consider a situation in which this is not the case, and point out its consequences.

In each of the various Figures to be used below, on the horizontal axis is measured the Health of A, and on the vertical axis the Health of B.\(^{10}\) For the moment we assume that this is measured in QALYs over some appropriate time horizon, using some appropriate measure of central tendency, and can be treated as cardinally measurable and interpersonally comparable. But whatever unit is chosen, it is used uniformly on both axes. The line FF represents the frontier of the health possibility set, representing those feasible combinations of health for A and B which are under consideration at some point in time. The purpose is to decide which is the optimal outcome according to each philosophical position. In each case the optimum will be denoted by X.

**No side-conditions: no maximand**

---

\(^9\) Although we have considered each theory as a separate entity, it is quite likely that an individual citizen will subscribe to more than one of them simultaneously. In such a case there will arise the need to establish equity-equity trade-offs as well as equity-efficiency trade-offs, but this is a complication which we have left to the imagination of the reader.

\(^{10}\) A and B may be any subgroups within the community whose relative health status is an object of policy concern. Thus, depending on the context, it may be rich and poor, or men and women, or black and white, or smokers and non-smokers, or northerners and southerners, or urban and rural, or those alive today and unborn generations. At the level of principle the issues concerning the social welfare function will be the same. But the health possibility frontier is likely to be very different according to which sub-groups are on the respective axes.
For completeness, this cell accommodates radical nihilism or moral scepticism, which claims that it is not possible to make sensible judgements about health outcomes being fair or just. Whatever happens happens, and we cannot judge whether it is good or bad or right or wrong.

No side conditions: linear maximand: weights equal

From a utilitarian standpoint, justice is ultimately a matter of maximising the sum total of human happiness.\textsuperscript{11} This principle embodies a kind of equality since it gives equal weight to each individual's happiness, in line with the Bentham's famous slogan: "each to count for one, none for more than one". Although there are many different brands of utilitarianism, we can identify three common features. First, consequentialism: things must be evaluated in terms of their consequences. Second, welfarism: consequences must be evaluated in terms of the welfare or utility of individual human beings. Third, sum-ranking: the overall evaluation must be based on the sum total of individual utilities in the relevant population. This latter principle means that the distribution of utilities between people is of no intrinsic concern; it also means that utilities must be cardinally measurable and interpersonally comparable.\textsuperscript{12} Different brands of utilitarianism can then be distinguished, among other things, according to how "utility" is understood (e.g. pleasures and pains versus desires or preferences) and how the unit of evaluation is defined (e.g. particular acts versus general rules or motives for acting).

Despite their lack of intrinsic concern with the interpersonal distribution of utilities, the eighteenth- and nineteenth-century fathers of utilitarianism, Bentham and Mill, were in fact passionate advocates of radical social reforms which would re-distribute income, health care and other utility-yielding goods from rich to poor. This is because utilitarianism yields a clear case for re-distribution of a good if one assumes diminishing marginal utility of that good, and if utility declines in the same way for everybody. Re-distribution is then bound to be a good thing on a utilitarian calculus, since the gain in happiness to the poor from one more unit of the good will be greater than the loss in happiness to the rich from one less unit of the good\textsuperscript{13}.

It is important to distinguish this classical utilitarian argument for equality of health (i.e. that greater equality of health generally increases the sum total of utility) from more fundamental egalitarian principles which assume that inequality in health is intrinsically bad. The classical utilitarian argument concerns the relationship between health and utility: it assumes that utility is diminishing in health.\textsuperscript{14} By contrast, more fundamental egalitarian principles focus attention on how individual utilities are aggregated to yield social welfare. Rather than simply adding utilities up, as classical utilitarianism does, these more fundamental egalitarian principles adopt a

\textsuperscript{11} Perhaps the best introduction to utilitarianism is Smart and Williams (1973). An excellent set of critical readings is Sen and Williams (1982).
\textsuperscript{12} Conventional Paretian welfare economics, because it works with ordinal utilities, rejects sum-ranking. Kenneth Arrow (1973) has called this position "ordinal utilitarianism".
\textsuperscript{13} Utilitarians also have a second general argument for equality, based on what Hare has called the "disutility of envy" (Hare 1993). This is the idea that individual utility may depend on the degree of inequality in society, both due to purely altruistic concerns for one’s fellow human beings, and to concerns about the crime and social disruption that may accompany gross inequalities.
\textsuperscript{14} It is more conventional, and perhaps more plausible, to assume diminishing marginal utility of commodities (such as consumer goods, or health care) rather than diminishing marginal utility of health. Furthermore, it could be argued that our QALY measure should be an all-inclusive utility measure, which already takes account of any diminishing marginal utility of health (see Chapter 34 by Dolan for further discussion of this "welfarist" argument).
social welfare function which gives greater weight to individuals with lower utility.

Despite its noble historical pedigree as a philosophy of social reform, utilitarianism has been subjected to a great deal of criticism in the philosophical literature in recent decades (Hampshire 1978). Each of the three key features of utilitarianism mentioned above have been criticised, with perhaps the fiercest criticisms being directed against consequentialism, on the grounds that the ends (i.e. the consequences) do not always justify the means. However, much of this criticism has been directed against utilitarianism as a system of personal morality, rather than a way of evaluating public policies and institutions. It has been argued, for instance, that utilitarianism recommends that one should become a cold and calculating person, who lacks integrity and dignity, breaks ties of personal affection to family and friends, and violates other common-sense ethical principles, in the pursuit of impersonal humanitarian ends (see, in particular, the essay by Bernard Williams in Smart and Williams, 1973). Yet while a close attachment to family, friends and personal projects may be virtues of personal morality, it seems reasonable to argue that matters of public policy properly require a more detached and impartial perspective which eschews personal favouritism of this kind (Goodin 1995).

[Figure 1 about here]

The conventional welfare economics approach is very close to the utilitarian position, being a simple maximising one as exemplified in Figure 1. Here a unit of health is regarded of equal social value no matter who gets it, which is denoted by welfare contours which are at 45° to each axis, implying that only the total amount of health is important, and its distribution between A and B is of no public policy interest. If the additional assumption is made that there is diminishing marginal utility from health, and that it diminishes in the same way for everybody, then we move on to Figure 3, which will be discussed later when we come to the relevant cell in the Table.

But this initial depiction actually departs from the currently conventional economic approach in two respects. Normally, what would be on the axes would be the utility enjoyed by each individual (bringing it even closer to the utilitarian position), not the health enjoyed by each individual. But for the time being we are restricting our attention to health and health alone as a utility-generator for each individual, and treating a unit of health as if it were a unit of utility. The second respect in which Figure 1 differs from what modern economists normally do is that, in order to fulfil the Pareitian restriction on the admissibility of statements about improvements in social welfare, the starting position becomes relevant. The reason for this is that a Pareitian would not accept that in Figure 1 the Health of A and the Health of B could be cardinally measurable and interpersonally comparable. To avoid these rather strong requirements, Pareto suggested that, when making judgements about the relative desirability of different social states, we should restrict ourselves to situations in which all that would be required would be for each person to say whether they were better off or worse off (according to whatever criteria they each separately chose to apply). Only those situations in which no-one was worse off and at least one person was better off should then be regarded as clear social improvements. The effect of this

---

15 One of the most influential critics of consequentialism was Isaiah Berlin. His famous essay "two concepts of liberty" (Berlin 1958) contains, among other things, a powerful critique of the idea that there must be a single true answer to moral questions about how good or bad particular consequences are for people and for society.

16 Utilitarians have responded to this charge by claiming that the utilitarian calculus can operate at the level of general rules for behaviour rather than particular acts. It may then turn out that the utility-maximising rules for behaviour are ones which respect common-sense ethical codes (to keep promises, to honour thy father and mother, and so on).
restriction is illustrated in Figure 1A, in which Z is the starting point, assumed to lie within the frontier. Pareto restricts admissible judgements about improvements in welfare to the quadrant lying to the north east of Z, in which Y would be the best attainable situation. In utilitarian terms, X is better still, but a Paretian can say nothing about a move from Z to X because B is worse off at X than at Z. For a Paretian to break out of this impasse, some way would have to be found for A to compensate B so that B would no longer be worse off than at Z, and A would still have some net gain left. The application of this "compensation principle" is not explored further here, because direct transfers of health between individuals are problematic, and the matter is better considered in a broader institutional context than that which we are currently supposing.

[Figure 1A (Pareto) about here]

No side-conditions: linear maximand: unequal weights

This would be a variant of the utilitarian position in which a unit of health would have a different social value according to who gets it, but the weights would be invariant with the amount of health a person has. It would represent a situation in which it might be argued that a particularly deserving class of people (e.g. war heroes) should always have priority over others when there are ways of improving people's health, no matter how much health they already have. The analytical representation of this notion simply requires the (straight-line) contours in Figure 1 to be at an angle greater than 45° to the axis of the more favoured group. This is the situation shown in Figure 2, where A is held to be the more deserving group.

[Figure 2 (principle of desert) about here]

No side-conditions: non-linear and smooth maximand: equal weights

The egalitarian goal of reducing inequalities in health has powerful appeal, and has motivated a large empirical literature on health inequalities (especially between social classes in developed countries, and between men and women in developing countries). As indicated above, it may derive from classical utilitarianism combined with an assumption of diminishing marginal utility of health. Alternatively it may derive from more fundamental objections to inequalities in health, irrespective of any effect on the sum total of health or utility. It seems unreasonable to insist on strict equality in health, however, since this might require a levelling-down in everyone's health towards that of the most unhealthy. So theories in this cell all accept a trade-off between equality and efficiency - the degree of curvature of the maximand reflecting the degree of aversion to inequality.\(^\text{17}\)

One such theory is the so-called “fair innings” argument, according to which there is some length of life (e.g. “three-score-years-and-ten”) which can be regarded as an ethical entitlement. Those who get less than this are entitled to feel unfairly treated, whereas those who get more than this have no cause to complain on equity grounds when they eventually die. In the present context this argument needs reformulating in terms of quality-adjusted life expectancy rather than just life expectancy, since a life spent disabled and in pain cannot be held to be equal to one of the same length spent in good health (Williams 1997). This fair innings argument would thus lead to priority being given to the young over the old on equity grounds. This goes beyond the implicit form of "utilitarian ageism" which follows from the fact that the elderly will,

\(^{17}\) At one polar extreme, we can think of zero aversion to inequality as reflecting the straight line maximand employed by classical utilitarianism. At the other polar extreme, maximum aversion to inequality would give rise to a kinked L-shaped maximand, of the kind to be discussed shortly.
generally speaking, have lower capacity to benefit from a given health intervention than the young.

The main criticism that has been levelled against the fair innings argument is that this strong form of "ageism" is incompatible with the duty of care that a civilised society owes to its elderly population. This objection is somewhat weakened, however, by the fact that the principle operates alongside a continuing desire to improve everyone’s health, and allows for a variable degree of aversion to inequality according to the mores of society. The uncomfortable conclusion, however, is that the more strongly one wishes to reduce inequalities (e.g. between social classes) in people’s lifetime experience of health, the more one will have to discriminate against the elderly and in favour of the young. This presents a further moral dilemma for all concerned.

A second theory that can be interpreted as falling within this cell is Sen's theory of capability (Sen 1980, 1993). On this theory, justice involves taking note of individual advantage as reflected by an objectively measurable index of people's capability to do valuable acts (e.g. to work, to play sport) and to reach valuable states of being (e.g. self-respect, good health). Aggregative considerations are assessed in the space of capability, as well as distributive concerns, and insofar as equality is involved as one of the competing objectives of justice, the concern is with equalising - or moving towards equalising - people’s advantages as measured by their respective capabilities. What goes on to the list of valuable functionings, and how the various weights are to be derived, are matters which Sen deliberately leaves open, to be determined by "reasoned agreement" according to the task in hand. We can interpret this theory as falling into this cell so long as (i) we focus on health and set aside other capabilities, and (ii) we allow trade-offs between equality and efficiency. This interpretation is not unreasonable, since Sen himself has recently argued that health is one of the most important indicators of well-being (Sen 1998).

The main criticism that has been levelled at Sen's theory is that it permits a high degree of subjective judgement on the part of the analyst in selecting the list of functionings and then determining how much weight to give each one. This strikes many liberal political theorists and economists as giving the analyst too much scope for heavy-handed paternalism (Sugden 1993). Sen's reply is that the capability approach is sufficiently general to be compatible with a wide range of specific conceptions of the good life and that "the need for selection and discrimination is neither an embarrassment, nor a unique difficulty, for the conceptualisation of functioning and capability" (Sen 1993).

Compared with the simple quasi-utilitarian situation depicted in Figure 1, these particular concerns about distributional justice can be represented by making two special assumptions: firstly that there is diminishing marginal social welfare from increasing the health of each individual, and secondly that this schedule is identical for both individuals. This would have the effect, shown in Figure 3, of making the welfare contours convex to the origin and symmetrical about the line OE (which is at 45° to each axis). In this case there is some trade-off between maximisation and equalisation, and the greater the weight given to equalisation relatively to maximisation, the greater will be the curvature of the social welfare contours.

One further difficulty with this interpretation is that it runs roughshod over the key distinction Sen makes between "functionings" (achieved outcomes) and "capabilities" (opportunity to achieve good outcomes). Sen argues that individual advantage should be assessed not simply as the bundle of valuable functionings (e.g. good health) the individual actually achieves, but as (an index of) the whole set of bundles he is capable of choosing from - his "capability set".

---

18 One further difficulty with this interpretation is that it runs roughshod over the key distinction Sen makes between "functionings" (achieved outcomes) and "capabilities" (opportunity to achieve good outcomes). Sen argues that individual advantage should be assessed not simply as the bundle of valuable functionings (e.g. good health) the individual actually achieves, but as (an index of) the whole set of bundles he is capable of choosing from - his "capability set".
According to theories of substantive equality of opportunity (LeGrand 1982, 1991, Arneson 1989, Cohen 1989, Roemer 1998), justice involves compensating people for any disadvantages they suffer through no fault of their own but not for disadvantages they suffer as a result of their own free choice. That is, everyone should have the same opportunity to obtain good things in life (e.g. health, wealth, positions of authority) but it should then be up to them to choose how they exercise that opportunity. It is important to distinguish this principle of substantive equality of opportunity from the much weaker principle of formal equality of opportunity, according to which people should have equal basic liberties and there should be no formal legal barriers against particular groups of people. Under formal equality of opportunity, however, people may suffer disadvantage through accidents of birth over which they have no choice - such as being born into a harmful social environment, or with disability, or with less natural talent than others.

Any theory of substantive equality of opportunity requires a theory of "free choice" to determine what counts as being freely chosen, and what counts as being beyond one's own voluntary control. Such a theory might tell us, for example, that people in managerial occupations generally have a higher degree of free choice about whether or not to adopt a healthy lifestyle than those in manual occupations. If so, equality of opportunity would then recommend that people in managerial occupations should receive less health care for any condition affected by their behaviour (and/or face greater taxes on unhealthy behaviours such as smoking) than those in manual occupations.

As before, a strict insistence on equality without any trade-off between equality and efficiency is unattractive, because it might require a levelling-down of opportunities for health (e.g. by withdrawing health care from those who currently enjoy higher-than-average opportunity for health). So equal opportunity theorists generally accept the need for balancing equality goals with efficiency goals.

If on such ethical grounds public policy takes a different view of the health of A relatively to the health of B, the welfare contours between the two are no longer symmetrical about the locus of equality (OE) in figure 3, but skewed in favour of one of them (say A). This is the situation depicted in Figure 4.

One criticism of the principle of equality of opportunity is that it is rather harsh and unforgiving to those who, for whatever reason, make decisions which they regret later in life. It amounts to justice without mercy. At a more practical level, another difficulty is that the available data may be too crude to measure adequately a concept as subtle and context-dependent as "free choice". So, in practice, the application of an equality of opportunity principle to public policy decisions will inevitably end up penalising some individuals who appear to have a high degree of free choice according to the available data, but who actually have suffered disadvantage through no fault of their own.

A final difficulty lies in restricting the principle of equality of opportunity to health alone. Arguably, one cannot have genuine equality of opportunity for health without also having equality of opportunity for all goods (LeGrand 1991, Culyer and Wagstaff 1993 p.445). This is because, if there is diminishing marginal utility of non-health goods (e.g. income), then those
who are disadvantaged in non-health terms will have to make a greater sacrifice of utility in order to attain the same level of health.

No side-conditions: non-linear and kinked maximand: equal weights

A strict egalitarian position would permit no concessions to health maximisation except as between alternatives that gave people equal levels of health. Strict equality at a higher level of health for both parties would be better than strict equality at a lower level of health for both parties. This is not to be confused with the argument that if we want to insure ourselves against being left in poor health when others enjoy better health, it would be prudent for us to insist on equality of health outcomes as the criterion to be employed in choosing between policy options. This justification for adopting the “maximin” principle would be the outcome of strong risk aversion rather than a strong desire for equity, where equity is held to be the moral imperative to maximise the level of health attained by the worst-off group.

[Figure 5 (maximin) about here]

The L-shaped indifference curves in Figure 5 represent this situation, in which improving the health of whichever group already has the greater amount of health does not register as a social improvement, but improving the health of the worse off person does (up to the point where equality is achieved). Thus moves towards greater equality improve welfare, as does maintaining equality at a higher level of health for both parties.

A variant of this maximin position may also be of some importance, because it appears, from some empirical survey work designed to elicit popularly held views about equity in health, that for some people public policy interest in reducing inequalities in health is nil until such inequalities reach a certain magnitude, at which point they dominate the situation and are regarded as intolerable. This threshold effect is depicted in Figure 5A, where close to the locus of equal health the welfare contours are as they were in Figure 1, but once inequality reaches a certain level the contours become parallel to the axes, as in Figure 5.

[Figure 5A (threshold) about here]

But it is also conceivable that people might hold the opposite view, namely that equality is the primary objective and its pursuit should only be moderated when the cost in terms of the overall health of the community becomes intolerable. This generates the situation depicted in Figure 5B, which shows a case where the chosen option is one where there is marked inequality, because the level of health that would have been enjoyed at the best feasible level of equal health (point Z) is too low a level of health to be acceptable.

[Figure 5B (on reverse threshold) about here]

No side-conditions: non-linear and kinked maximand: unequal weights.

In the preceding case the characteristics of the two parties played no role in determining the optimal outcome, but it might be the case that considerations such as those discussed earlier in relation to Figure 4 might come into play. This was the situation in which it might be held that one party is more deserving of special consideration by society because they were not in any way responsible for their own ill health. A similar argument might be put forward when it is

---

19 It is also possible that these contours are curved as in Figure 3, though this case is not shown here.
the fact of inequality in health that is the focus of ethical concern, and not its cause, and when the welfare contours brook no compromise with health maximisation except when choosing between options which satisfy the equity objective.

One possible instance of such a position is the notion that health policies should favour those whose current situation is worst, even though they are not the people for whom the greatest improvements in health are possible. In this scenario, “need for health” is the driving force, and it is argued that the objective should be to choose those policy options that would most quickly bring those people into equality with the others, but without reducing the health of the better off. This means that the starting point becomes relevant, as it was in the Paretian case depicted in Figure 1A, but now the welfare contours have become L shaped, indicating that one particular inequality-reducing path is preferred to all others. This is the situation depicted in Figure 6, where Z is the starting point, at which the health of A is worse than the health of B. The L-shaped welfare contours are no longer symmetrical about the locus of points of complete equality, OE, but about some locus such as ZY, where Y is the “target” level of health distribution to which policy aspires. This represents a policy which seeks to reduce inequalities by sharing health improvements between A and B in a way that systematically favours A. In the case shown, where ZY is a straight line, this is in a constant ratio. It implies according a greater weight to health improvements for A compared with the weight given to the same improvement for B.20

[FIGURE 6 (distribution according to health deficit) about here]

One side-condition: no maximand

We have found no notion of distributive justice which fits into this cell, although such a notion is quite plausible. It would simply assert that no situation could be considered equitable which did not achieve some decent minimum level of health for people21. But lacking a maximand, there would be no means of discriminating between situations which met this condition, or of deciding which is the better situation amongst those which fail to meet the side-condition. It would be a strictly dichotomous judgement, equitable or not equitable, with no degrees of differentiation within either set. It is the situation represented in Figure 7, where there may exist many health possibilities which satisfy the side condition, but in the absence of a maximand there is no way of choosing between them. This is equivalent to shifting the origin from O to O’, and ruling out the shaded area to the southwest of CO’C. But it is also conceivable that the minimum requirement may be set so high that it is completely unattainable, or, in a particular policy context, that it has to be adjusted somehow so that it just touches FF at a single point (its intersection with OE). In this latter case policy would be being driven by practicality, and in the absence of any explicit optimising criterion, the adjustment of the “decent minimum”

20 This is not unlike the ethical position advocated by Murray and Lopez (1996) who argue that everyone should be assumed to have a life expectancy at birth of about 80 years so as to prevent those with short life expectancy being discriminated against when calculating the benefits of health care interventions. This is equivalent to assigning a greater weight to each additional year of life gained by those who actually get only a few, than is assigned to each year gained by those who actually get a lot. If the weights are inversely proportional to the relative distance of the two parties to the common target, it would generate a straight line such as ZY

21 The notion of a decent minimum of health is to be carefully distinguished from the notion of a decent minimum of health care, as proposed for example by Fried (1976). The latter principle would act as a constraint on health outcomes, requiring a minimum package of health care services irrespective of the health gains that might be enjoyed by redeploying resources elsewhere.
to whatever is actually attainable would presumably be guided by political expediency rather than by any ethic concerning distributive justice.

One side-condition: linear maximand: equal weights

A utilitarian position, modified by imposing a side constraint that people must have a decent minimum level of health, would fall into this cell of our classification. It is a plausible position, which might generate three rather different outcomes. The first would arise if the side condition were not satisfied. The second situation would arise if the side condition were fulfilled, but it prevented the health-maximising situation from being attained (i.e. no matter how big the sacrifice imposed on B, A's needs must be satisfied). The third situation would be where the side condition was easily satisfied and the outcome is no different from what it would have been had a simple utilitarian position been adopted.

One side-condition: other permutations

In principle, each of the other maximands can be combined with a side condition. We could also have more than one side condition. In this case, the side conditions can either be satisfied simultaneously or sequentially. As before, the side conditions can in principle be combined with any of the maximands. However, we have found no published philosophical work which examines these complex permutations.

THE OPPORTUNITY SET ONCE MORE

So far the argument has been conducted with a feasibility set that has the normal properties assumed in economics. But in the health field there is an important set of circumstances which require the feasibility frontier to take on a different shape. This occurs when the health of one person is directly affected by the health of another person (e.g. with infectious or contagious diseases, or in mental illness where a person may become a threat to other people’s safety). In such a case the health of (say) B can only be improved if the health of A is improved.

A similar diagram is used in Mishan (1977) and Atkinson and Stiglitz (1980).
NEEDS-BASED THEORIES

Several philosophers have argued that health care should be distributed according to need (including Bernard Williams 1962, Harris 1988, Lockwood 1988, Waltzer 1983). Typically, these philosophers do not argue that all goods and services should be distributed according to need; rather, they argue that health care is a special kind of good which requires a special principle of distribution. Tobin (1970) calls this idea "specific egalitarianism". Waltzer (1983) develops this idea into a theory of "spheres of justice", according to which different goods require different principles of distribution according to the nature of the good (consumer goods, for instance, should be distributed according to market forces; higher degrees according to academic merit; health care according to need).

One difficulty with needs-based theories lies in drawing the line between goods which should be distributed according to need, and those which should be distributed according to market forces. Should goods other than health care also be distributed according to need (e.g. nutrition, shelter, transport, employment)? For that matter, should some forms of health care be distributed according to market forces (e.g. cosmetic surgery)? A second difficulty lies in defining what "need" means. Different definitions can recommend very different distributions of health care. We can identify at least three possibilities: (i) the person's initial level of ill-health, (ii) the person's capacity to benefit from health care, and (iii) the expenditure required to equalise health. Definition (i) leads to the position shown in Figure 6. Definition (ii) leads to the simple utilitarian principle of health maximisation shown in Figure 1 (since care is given to those who benefit most). Definition (iii) leads to a side constraint requiring strict equality of health (since care is given so as to equalise health). For a more detailed analysis of these and other definitions of need, the reader is referred to Culyer and Wagstaff (1993).

IV THEORIES WHICH ETHICALLY CONSTRAIN THE OPPORTUNITY SET

INTRODUCTORY COMMENTS

There is a strong philosophical tradition which approaches issues of distributive justice in a totally different manner. It is argued that there are certain desiderata about the way in which public policy decisions are made which, if observed, will ensure that the outcomes will be just, and no further criteria concerning fairness need to be applied to them. To incorporate this mode of thinking into conventional economic analysis we need to think of it as restricting the feasible opportunity set to a narrower "ethically acceptable" set. Thus, in Figure 10 the original frontier FF has been shrunk illustratively to F'F'.

But this "illustration" requires some further commentary. First of all, there is no guarantee that the specified desiderata will leave any feasible outcome "ethically acceptable" at all. Conflicts between ethical principles may arise which require one or other principle to give
way. Collectively the desiderata may indeed be vulnerable to an "impossibility theorem" which says that something will have to be relaxed for a feasible acceptable outcome to emerge. To choose which desideratum to relax requires some higher level optimisation criterion to be added, so that each constraint can be "shadow priced" in relation to it. But such theories do not typically have any such "maximand", but merely assert the intrinsic value of each constraint. However, suppose that as a result of some further consideration, the key constraint were adjusted (downwards) to make just one feasible outcome ethically acceptable, then Figure 10 would need to be redrawn such that \( FF' \) became a single point somewhere within the original frontier FF.

But not all theories which restrict the opportunity set are cast in these strict terms. Some are hybrids which, although initially working their way through a series of conditions which have to be satisfied lexicographically, allow whatever is left to be subjected to a maximising rule. In Rawls' theory, for instance, the first task is to maximise the liberty enjoyed by each individual, subject to that same amount of liberty being available to all. This maximisation problem takes lexical priority over the second one, which is to maximise the welfare of the worst off individual. Thus, once the maximal "equality of liberty" condition has been satisfied, outcomes are to be ordered according to the extent to which they improve the lot of the most disadvantaged in terms of an index of "primary goods" such as income, authority and respect (Rawls 1982).

In what follows we shall first of all discuss theories which can be interpreted as imposing ethical constraints upon the health possibility set, but offer no maximand. The bulk of philosophical thinking in this field seems to fall into this category. Then we shall consider theories which do have such a maximand. As will be evident from the table this appears to be sparsely populated territory. This may well be because we have not come across any such theories in our search of the philosophical literature. It is an interesting (and perhaps even a productive) intellectual exercise (which we leave to the reader) to formulate theories of distributive justice which would fit into each of these empty cells.

THEORIES WITH NO MAXIMAND

1. Libertarianism:

According to libertarians (e.g. Nozick 1974, Locke 1967), justice is entirely a matter of the enforcement of private property rights, including the right to the fruits of your own labour. The role of the state should be to provide law and order, but no more than that. The "minimal" or "night-watchman" state should safeguard property, deter force or fraud in private contracts, and force criminals to pay adequate financial compensation to their victims. But the state should not raise taxes or interfere with individual liberty for any other purpose - not to regulate markets or to provide public goods, nor to re-distribute income and health from the fortunate to the less fortunate. So libertarians reject the notion that there should be governmental agencies with direct responsibility for efficiency and equity in the provision of health. Instead, they advise that state involvement in health care should cease, and be replaced by private health care insurance and private charity for the poor and the disabled.

For libertarians, it makes no sense to enquire about the justice of a particular end-state distribution of benefits and burdens (such as health outcomes). Instead, what matters is the

---

23 In a similar vein, ethical constraints may not take absolute and unconditional priority over efficiency considerations: the constraint may be relaxed once the "shadow price" of upholding it becomes too high.
historical pattern of events. So long as the distribution of health outcomes results from a process of free private contracts between consenting adults, then it is a just distribution. Gray (1977) puts it thus: "Attempts to impose any other principle on the free exchanges of free men involve imposing upon them a hierarchy of needs and merits, about which no consensus exists in our society and which there is no reason to suppose can be achieved."

An important feature of libertarianism is that rights are seen as absolute ethical constraints, and that no trade-offs are permitted under any conditions. In particular, the state should not violate minor rights in order to prevent major rights from being violated (e.g. for the police to steal private documents in order to prevent suspected terrorists from blowing people up). According to libertarianism, the state has no business violating one person's rights for the sake of someone else's rights, which would amount to what Nozick (1974) has called a "utilitarianism of rights".

2. Participatory democracy:

Another important strand of thinking about justice that fits into this cell is what might be called the "participatory democracy" school of thought. This is the idea that justice is entirely a matter of having a suitably free and fair process of democratic dialogue, in which all citizens participate on an equal footing. So long as this process of dialogue is in place, the resulting health outcomes must be just. A notable exponent of this dialogue theory of justice in the health care literature is Rudolf Klein (1996); an exponent in the philosophy literature is Ackerman (1980). Instead of inviolable rights to private property, individuals are held to have inviolable rights to participate in public dialogue.

It is important to distinguish the principle of (actual) dialogue as a substantive theory of justice from the advocacy of (hypothetical) dialogue as a means for choosing between rival theories of justice. As a substantive theory of justice, the dialogue principle can be criticised as being too vague to give guidance to policy-makers and health care managers. Without further specification, it gives no guidance as to what procedure of dialogue is desirable; or how and why one procedure of dialogue is to be chosen over another. Nor, once all the talking is done, does it help the person who is charged with making the final decision.

Dialogue as a means of choosing between rival theories of justice has been advocated most notably by Habermas (1984) in his theory of communicative action. Habermas argues that the correct methodology for finding a theory of justice is to ask what rational individuals would agree to after a process of dialogue in an "ideal speech situation" - i.e. under conditions of free and uncoerced discussion. In principle, however, any substantive theory of justice might be agreed to in the ideal speech situation (including utilitarianism, libertarianism, and so on), so the task of analysing them all, in order to inform such a dialogue, remains.

3. Contractarianism:

Contractarianism, or social contract theory, is a broad tradition of political thought based on the idea of a hypothetical agreement between rational individuals about what principles of justice should govern their social interactions. The notion of a hypothetical social contract is amenable to economic analysis, although this approach departs markedly from the conventional welfare economic approach as described in section II. Contractarian analysis often leads to theories of justice which specify ethical constraints rather than maximands (e.g. Buchanan 1975, Hayek 1976, Sugden 1986). In principle, however, depending on the precise bargaining situation or "original position" the parties to the social contract find themselves in, the contracting parties might come to agree upon any substantive theory of justice, including ones with maximands.
Rawls' (1971) theory of justice, which specifies that the contracting parties are operating from behind a "veil of ignorance" about their personal characteristics and position in society, is a contractarian analysis which leads to a theory with both ethical constraints and a maximand. Harsanyi's (1955) theory of utilitarianism, which argues that rational individuals from behind a "veil of ignorance" would choose to maximise lifetime expected utility, is a contractarian analysis which leads to a theory with a maximand but no ethical constraints. Contractarianism is thus perhaps best seen as a general methodology for choosing between theories of justice, rather than as any one particular theory of justice in its own right.

4. No-envy principles and equality of resources:

Some economists have explored the idea that a fair situation is one in which no-one envies anyone else (Varian 1974, Baumol 1986). The idea is that no-one should prefer to be in any one else's shoes, taking into account not only their economic situation (e.g. what goods they consume and what work they do) but all relevant aspects of their personal circumstances (e.g. what disabilities they suffer from). The trouble with all "no-envy" principles, however, is that they require a compensation principle to say anything useful about real-world situations in which the ideal situation of no-envy cannot be realised. Without a compensation principle, we can say nothing about how reductions in envy for some can be traded-off against increases in envy for others.

One philosopher who has tried to develop a compensation principle within the no-envy framework is Dworkin. Dworkin (1981a,b) uses the philosophical device of imagining a hypothetical Walrasian insurance market for individuals who have not yet been born. Compensation is then defined as the amount of insurance which (unborn) people with equal cash endowments would take out against the possibility of being born disabled or lacking in talent. The motivation behind this hypothetical device is that it allows us to set a market value not only on transferable "external" resources (goods and services) but also on non-transferable "internal" resources (talent, intelligence, health). Equal cash endowments in this hypothetical situation thus embodies a no-envy principle which takes into account both external and internal resources.

Dworkin's central thesis is that justice requires that people should be compensated if they have disabilities and lack of talent but not if they have expensive tastes or foolish ambitions. He argues that people have a degree of responsibility for cultivating their own preferences, and so should not be compensated for having expensive tastes and over-weaning ambitions. In the context of health care, for example, it can be argued that preferences for frivolous beauty treatments, held by vain and spoilt patients, should not attract resources away from less demanding patients who are in genuine need of health care. More generally, Dworkin's idea is that people should be compensated for disabilities, which are an involuntary accident of birth, but they should not be compensated for expensive tastes, which are voluntarily cultivated. However, this way of drawing the line between compensatable and non-compensatable factors has attracted criticism from Cohen (1989) and Arneson (1989). These theorists point out that some expensive preferences may be relatively involuntary (e.g. a taste for angling which suddenly becomes expensive due to pollution) and that some disabilities may be relatively voluntary (e.g. becoming disabled due to a sports injury). They argue that the degree of free choice should determine the degree of compensation - which then leads back to an equality of opportunity theory such as we considered earlier.

A second criticism of Dworkin's theory is that it makes no allowance for efficiency goals and insists upon full equality of resources. This is unattractive, since it might mean a levelling-

---

24 Roemer (1996) contains a critical summary of the details of this compensation proposal.
Dworkin replies that this consequence only appears unattractive if one adopts an overly materialistic conception of individual well-being. Dworkin adopts what he calls a "challenge" conception of well-being, according to which individual well-being is entirely a matter of reacting appropriately to challenges in appropriate circumstances - i.e. living well in a good society (Dworkin 1988). He contrasts this with an "impact" conception of well-being, according to which well-being is entirely a matter of achieving appropriate final ends (for example, pleasure, health or whatever). On Dworkin's "challenge" conception, well-being is largely a matter of individuals living their own lives well or badly. All the state can do is to get the circumstances right - which, in Dworkin's view, means equality of resources for all, and not necessarily more resources for all.

5. Equality of access to health care:

A number of economists have proposed the principle that access to health care should be the same for everyone - or for everyone with equal need (e.g. Mooney 1983, Mooney et al. 1991, Olsen and Rogers 1991). This principle focuses exclusively on health care, ignoring all other possible ways of improving health, and so differs markedly from the principle of equality of opportunity for health outcomes. The principle of equal access to health care places ethical constraints on the health possibility set: it rules out all attainable health outcomes that require unequal access to health care. By contrast, the principle of equal opportunity for health leaves the health opportunity set unconstrained, but generates an asymmetrical maximand (as we saw earlier, in Fig 4).

The main criticism of the equal access principle is precisely that it focuses too narrowly on the commodity of health care, rather than looking at what health care can actually do for people. A second difficulty is that it is not clear how "access" should be defined. We can identify at least four operationally different definitions: (i) access as the utilisation of health care; (ii) access as the money and time costs incurred in receiving health care (Mooney 1983), (iii) access as the maximum attainable consumption of health care (Olsen and Rogers 1991), and (iv) access as the forgone utility cost of obtaining health care (LeGrand 1982). The reader is referred to Culyer and Wagstaff (1993) for a discussion of the pros and cons of each definition.

6. Medical ethics and the rule of rescue:

So far, we have discussed ethical constraints that have been proposed in the context of decisions about groups, rather than particular individuals, since that is the focus of this Chapter. However, ethical rules intended to govern decisions about particular individuals and patients (e.g. do no harm, respect autonomy, tell the truth, and so on) will often have consequences for resource allocation at the group level, and to this extent can be thought of as further ethical constraints on the health possibility set. There is a large medical ethics literature which proposes ethical principles for individual-level clinical decisions. One such principle has been termed the "rule of rescue" (Hadorn 1991). This is the idea that each individual, and society as a whole, has an ethical duty to do everything possible to help those in immediate life-threatening distress. In some cases, the rule of rescue is merely a common-sense rule of thumb which moves us closer to the health possibility frontier - for instance, it may sometimes be efficient to treat urgent cases at once, without immediate regard to cost, because a stitch in time saves nine. Or, there may be more convoluted justifications for this rule in terms of indirect or long-term consequences for health - for instance, people may gain greater peace of mind if they know that those in distress will always be rescued. However, it is hard to give a rational justification for the rule of rescue insofar as it conflicts with other principles of justice. It seems irrational to

---

25 Good introductions to the medical ethics literature are Glover (1997) and Gillon (1986).
devote resources to people who happen to be in immediate distress if it is true that other people with a greater claim on those resources (e.g. those who need them more, or who are more disadvantaged) will lose out as a result. In defence of the rule of rescue in such cases, however, it can be replied that some ethical rules are such deeply embedded social conventions that they should be followed unless the consequences of doing so are clearly disastrous.

7. General comments on these six theories:

In each of these six cases the implications for the analytical apparatus are the same and are shown in Figure 10. Of the options that were in the health possibility set when it was unconstrained by notions of equity or justice, some will have to be removed. But since the principles for deciding which ones are inadmissible do not refer to anything in the health outcome space, these deletions will follow no systematic pattern in that space. The options that are ruled out may only be those which lie well inside the frontier, but we have assumed that some will be on the frontier, so that the frontier shrinks (otherwise these theories would leave actual policy choices unaffected). But their main problem from an economist's viewpoint is that they offer no maximand, so no guidance is offered on which of the admissible options should be preferred, or on which of the “unjust” options would be preferred in a second-best situation where something has to be done (since maintaining the status quo in policy terms should be amongst the options). Instead of optimising subject to constraints, what we have here are simply the constraints.

THEORIES WITH A MAXIMAND

1. Equality of opportunity for health using financial means only:

It can be argued that a more equal distribution of health is a desirable maximand, but that certain means of achieving this (e.g. via forcible redistribution of kidneys or other body parts) are ethically unacceptable. A hybrid theory of this kind was proposed in LeGrand (1987), which argues for equality of opportunity for health as a maximand but then imposes an ethical constraint against using the health care delivery system to achieve this goal. Equality of opportunity for health requires discrimination against those who do not fully exercise their opportunity for health (e.g. against smokers from managerial classes). LeGrand (1987) argues, however, that this discrimination should be done only via health care financing (e.g. consumption taxes, insurance premiums, user chargers) and not via the delivery of health care. This theory leaves in place an equal opportunity maximand, but imposes the ethical constraint against health possibilities which require discrimination in the delivery of health care.

The difficulty with this kind of theory is that it is not clear how, or on what basis, the line should be drawn between acceptable and unacceptable means of redistributing health. In particular, the argument cannot be that financial redistribution is acceptable because it only involves redistributing non-health outcomes (e.g. consumer pleasures) which are less morally significant than health outcomes. This argument is false to the extent that (a) income and health are linked, and (b) non-health outcomes are morally significant. Furthermore, it is an argument for restricting the principle of equality of opportunity to non-health outcomes only, and employing some other maximand for health.

2. Rawlsian theory:

According to Rawls' theory of justice (Rawls 1971), there are two principles of justice. First, basic liberties are to be distributed equally and at the maximum level that is compatible with everyone else enjoying the same level. Second, social and economic inequalities are to be
arranged so that they are to the greatest benefit of the least advantaged members of society. Basic liberties include the right to vote and to be eligible for public office, the right to hold property, freedom of speech and assembly, and freedom from arbitrary arrest. Social and economic inequalities are to be measured by an index of "primary goods" that (i) are all-purpose means that every rational person needs to pursue their own ends and that (ii) are distributed by society rather than nature. This includes basic liberties, income and wealth, positions of responsibility, and self-respect, but does not include "natural goods" such as health, intelligence and imagination. These principles are to be satisfied sequentially, the first taking absolute lexicographic priority over the second. This means, in particular, that people are not allowed to trade off basic liberties for greater income (or, indeed, greater health).

Figure 11 illustrates our (re-)interpretation of the Rawlsian position as applied to health. We have interpreted Rawls first principle - the "liberty principle" - as an ethical constraint on the health possibility set. This is because it will rule out any attainable distributions of health which require inequality in basic liberties. We have interpreted the second principle - the "difference principle" - as an L-shaped maximand in the space of health. However, we would like to emphasise that this interpretation of the difference principle is explicitly rejected by Rawls (1982), who insists that the principle should operate in the space of primary goods rather than health. One reason is that health is distributed by nature as much as by society. A second reason is that good health is an end in itself as much as an all-purpose means that every rational person needs for pursuing their own ends. The third reason is that, as Arrow (1973) has pointed out, a strict application of the principle of maximin health would require excessive resources going to those with extreme health care needs, perhaps to the extent at which others are reduced to subsistence levels of income. For these reasons, Rawls has explicitly stated that health is not a primary good and that he restricts his basic theory of justice to "normal, active and fully co-operating members of society" (Rawls 1982). As Daniels (1985) puts it, "In effect, there is no distributive theory for health care because no one is sick!" [italics in original].

In his more recent work ("Political Liberalism"), Rawls (1993) mentions one way in which his basic theory of justice might be extended to incorporate the possibility of ill-health. Following Daniels (1985) he suggests that, before the difference principle should come into effect, all individuals should be brought up to the minimum level of health required for them to be "normally functioning" members of society. This moves us to a different cell in our Table, involving a side condition on health outcomes (i.e. a minimum level of health). If we keep the same maximand as before (i.e. maximin health), we move to Figure 12.

---

26 Other authors have re-interpreted Rawls in this way, notably Green (1976) who argues that all basic needs (including housing, education, and so on) should count as primary goods.

27 This second reason is particularly important if our measure of health incorporates quality of life as well as length of life, and hence requires investigation into the subjective utility values that people set on differing levels of morbidity. One of the key motivations for Rawls' use of primary goods as the index of individual advantage, rather than utility or welfare, was to avoid his theory having "to examine citizen's psychological attitudes or their comparative levels of well-being" (Rawls 1982). Rawls argues that it is not possible to find a cardinal and interpersonally comparable measure of the quality of life that is universally acceptable, on the grounds that in a pluralist society people can and should have irreconcilably differing views about what constitutes a good life.
FITTING POPULATION CHANGE INTO THE FRAMEWORK

The issue of population change raises special difficulties which are not easy to handle within the framework we have outlined in this Chapter. If population size is treated as exogenously given, then health inequalities between born and unborn generations can be dealt with in the same way as those between any other population subgroups A and B. The difficulties arise when policy decisions influence population size in ways that we judge to be ethically significant. In particular, the use of an average measure of population health by our framework already presupposes a particular (and controversial) theory of optimal population. It implies that the state should encourage the birth of people who would have above-average health, and discourage the birth of people who would have below-average health.

Parfit (1984) and Broome (1996) have investigated the issue of optimal population size, and have shown that it leads to problems even when our framework is modified in various ways. One obvious modification would be to take the sum total of health for the population subgroup, rather than average health, as the measure of health on each axis. However, this leads to what Parfit (1984) calls the "repugnant conclusion", that a sufficiently large population of people living in (equal) pain and misery will turn out to be better than a smaller, healthier population. This is because adding people whose health is on average the same as that of the existing population will always increase the sum total of health without changing the degree of inequality. To avoid this conclusion, a further modification would be to subtract a "critical level" of health from each individual's health score before taking the sum total (Blackorby, Bossert and Donaldson 1995). However, this critical level must be set fairly high to avoid the "repugnant conclusion" that it is optimal for there to be a large population living just above the critical level of health. Yet this has potentially controversial implications, since critical-level theory recommends that the state should discourage the birth of those who would be below the critical level.

So it remains to be seen whether our framework can satisfactorily accommodate the issue of population change with further modifications, or whether this issue requires a rather different ethical framework.

V HOW IS HEALTH TO BE MEASURED?

Up until now we have assumed that a person's health is measured in QALYs over some appropriate time period. But in the context of any discussion of equity in health, other possibilities have been advocated, so this now becomes a matter of direct concern, and of sharp difference of opinion. In some contexts it is a person's current health state, in terms of an immediate threat to life (life expectancy), that is taken to be the dominant ethical concern ("the rule of rescue", see Hadorn 1991). In others it is the person's current level of suffering or disability that dominates (Nord 1993). In yet others it is the size of the health gain that is relevant rather than the resulting level of health (e.g. the idea that it is better to give a small gain to many than large gains to a few). For some (e.g. Harris 1988) the rest of a person's life is taken as the unit, and regarded as equal between people irrespective of its expected length or quality, conditional only on whether the person wishes to go on living. And for the "fair innings" argument it is a person's whole lifetime experience of health that is taken to be the valid unit of comparison (measured in terms of life years [in Murray and Lopez's 1996 DALY approach] or in terms of quality-adjusted life years [in Williams's 1997 approach]).

Thus each of the philosophical approaches outlined above could be further subdivided
according to which concept of health it employs when comparing the plight of A and B. Each such concept will generate different conclusions for policy. This is true even with the simple maximisation rule depicted in Figure 1. Maximising life expectancy is not the same as minimising suffering or disability. The implications of maximising health gains will be affected by how these are measured, and if "the rest of a person's life" is the maximand, only immediate survival (i.e. "saving a life") will count as a health gain. And the "fair innings" approach differs from all the others in taking people's past health history into account as well as their actual present health state and their expected future health prospects.

VI THE ECONOMIST'S ROLE

To monitor states of the world, in order to determine the extent to which they satisfy the postulated criteria for equity in health, requires observation. In the case of "pure" procedural theories which restrict the set of options on ethical grounds, the empirical task becomes that of determining whether or not the postulated conditions are met by each potential candidate (including the status quo). In general they will not be met, which inevitably leads to the conclusion that the distribution of health is generally not equitable. The implied solution to this problem is some radical reorganisation of society so that the conditions are in fact met, which typically will be such a vastly complex and socially costly enterprise that it cannot possibly be undertaken in one fell swoop. Moreover, it is usually difficult to decide where to start, because trade-offs are not permitted, so no guidance is offered by which to judge, on a piecemeal basis, where the greatest progress might be made at least cost. Generally speaking, therefore, economists have fought shy of becoming engaged in the evaluation of such reform processes, since until their advocates accept the possibility of trade-offs we have little or nothing to contribute from within our discipline, except to remind them of the ubiquitous existence of opportunity costs.

Moving on to those theories which have side-conditions on acceptable outcomes but no maximand (as in Figure 7), here again economic analysis has no role to play, because the purpose of policy analysis here is simply to establish whether there are any options that would generate outcomes that lie in the acceptable zone to the northeast of O'. Since there are no (explicit) criteria for choice apart from satisfying the side-conditions, there is nothing further to be done. This is equally true whether there are any acceptable options or not. As before, we are offered no guidance as to what best to do to move from wherever we are to some point in the acceptable zone (it does not matter whose health we sacrifice, or by how much, in order to get there). Again, the economist's principal role seems to be to draw attention to this hiatus in the formulation of the policy problem.

This brings us finally to those theories which have a role for a maximand, with or without prior side-conditions in the outcome space. Here there will be a role for economic analysis to help enlighten the public policy process. In the simplest case, where the satisfaction of an equity norm concerning the distribution of health may entail some sacrifice in the total amount of health available to the community as a whole, in principle the matter is a relatively simple one of estimating an equity-efficiency trade-off for health. As will soon become evident, however, it is not such a simple matter in practice, since estimating the properties of a social welfare function for health is almost as difficult as estimating one for social welfare generally. Where the equity...
norm is cast in terms of side-conditions upon acceptable outcomes, then for those side conditions that are binding the role of economic analysis lies in calculating the shadow price of the constraint at the point where it binds. This tells us the cost, in terms of the maximand, of being where the constraint holds us, as compared with being wherever we would otherwise have chosen to be in the outcome space (e.g., in Figure 8B above, the difference between the aggregate health of the community at point Y and the aggregate health of the community at point X). The community can then decide whether observing this ethical constraint is worth this sacrifice, or whether it wishes to moderate it somewhat, or make it even more stringent. Thus the "shadow price" becomes an input into an equity-efficiency trade-off consideration. In these cases an analyst needs to note which health concept is the relevant one for the theory of justice under consideration, and how groups A and B are defined. Each theory needs to be scrutinised carefully for its chosen "target" in the equity stakes.

VII ESTIMATING EQUITY-EFFICIENCY TRADE-OFFS IN HEALTH

If the nature and implications of these various equity principles are to be clarified in a policy-relevant way, it is necessary to quantify what might otherwise merely remain vaguely appealing but ambiguous slogans. Only with some quantification will it be possible to convert them into criteria that can be applied in a consistent manner, and with a reasonable chance of checking on performance (i.e. holding people accountable). Quantification thus has the potential for clarification, for performance measurement, for accountability, and for policy formulation, analysis and reappraisal. The quest for more quantification of equity issues is worth pursuing on these grounds alone, despite the hostility that that it is likely to engender from those who mistakenly equate greater precision with lack of humanity. This seems to be another manifestation of the difficulties some people have in feeling sympathy for large numbers of people whose fate is represented by the relevant statistics, when they would respond quite readily if these same facts were presented to them as a series of individual case studies.

Generally speaking, the work of economists in seeking greater quantification in this field falls into two classes: that which addresses equity-efficiency trade-offs in the distribution of health explicitly in a quantitative manner, but currently lack the empirical data with which to support the assumed numerical relationship; and that which attempts to estimate such trade-offs empirically using questionnaire methods, but without having an explicit theory of justice into which to insert and interpret the findings. It is a rather unsatisfactory state of affairs.

Williams' (1997) exploration of the "fair innings" argument is an example of the former type of enterprise. It starts from the commonplace observation that people who die young are widely regarded as not having had a "fair innings" compared with those who live to a ripe old age. Thus if one adopted a person's whole lifetime as the appropriate period of observation, and took something approaching quality-adjusted-life-expectancy-at-birth in that community as the "norm", then someone enjoying a poorer prospect than this would be entitled to feel unjustly treated, and deserving of more effort on the part of the community to improve his prospects than should be devoted to someone whose prospects are better than this norm. The exact magnitude of such "equity weights" will depend on how averse that community is to inequalities in people's lifetime experience of health. Going back to Figure 3 (and interpreting health in terms of whole lifetime health prospects), the greater the community's aversion to such inequalities between A and B, the stronger will be the curvature of each social welfare contour. The greater the curvature, the greater the sacrifice in total health that will be acceptable in order to improve the distribution of health. Where no such sacrifice is acceptable, we are back with Figure 1. Where having an equal distribution is all that counts, anywhere on OE is preferable to anywhere off it. Williams assumes that the general situation will be as in Figure 3, and uses hypothetical data about inequality aversion to generate equity weights between social classes, and between old and
young within social classes, in the UK, that would be consistent with the "fair innings" approach to equity in health. He also speculates that different weights might be relevant for different subgroups within the community, since people's aversion to inequality might depend on who the inequalities were between, and what had caused them. This last consideration would make Figure 4 the appropriate analytical framework.

Murray and Lopez (1996) are also concerned to reduce inequalities in lifetime health, but they concentrate on life expectancy rather than on quality-adjusted life expectancy. Their objective is to reduce inequalities in the burden of disease between different regions of the world. They generate implied equity weights in a rather more indirect way. In essence, they require that when calculating the distribution of the global burden of a disease, it should be assumed that in the absence of that disease everyone would enjoy the age-specific life expectancy of a hypothetical population whose life expectancy at birth is 80 years for males and 82.5 years for females. In the simplest terms, this means that although in a poor country someone's actual life expectancy at some age might be 30 years, it will be assumed to be much greater, say twice as great, thereby implicitly weighting each year at twice its actual amount (60/30). In a richer country, where the person's actual life expectancy at that same age would have been (say) 40 years, the implied weight is 60/40 or 1.5, and so on. There has been no attempt to test empirically whether this view of equity in health has any widespread support, or whether the choice of 80 or 82.5 years of life expectancy at birth as the norm for men and women respectively, is the one that best reflects in quantitative terms the desired equity-efficiency trade-off. It is the situation represented in Figure 6 earlier.

Turning next to the published studies that have used questionnaire methods to obtain empirical estimates of equity-efficiency trade-offs, we find disappointingly little evidence that can be usefully inserted into theoretical frameworks of the kind discussed above. Although a good deal of progress has been made in identifying key factors and key regularities in people's preferences for equity in health, very few studies have actually attempted to infer numerical parameters from their results that could be slotted into theoretical models of equity-efficiency trade-offs. In reviewing this work, we first describe the questionnaire methods that have been used and then turn to the main findings.

In principle, standard health valuation methods such as "standard gamble" or "time-trade-off" (described in Chapter 34) could be converted into methods for eliciting equity-efficiency trade-offs. In effect, people could be asked how much of their own health they would be willing to sacrifice in order to achieve equity in the distribution of health. To our knowledge, however, no study has used this personalised approach to eliciting equity trade-offs. Instead, most of the studies reviewed below have asked respondents to adopt the impersonal perspective of a social decision-maker who has to choose between alternative distributions of other people's health. A third approach, which lies somewhere between the personal and impersonal approaches, asks people to base their answers on what they themselves would prefer were they to face the (known or unknown) probability of being the individuals they are asked to choose between. In other words, respondents would be required to make personal choices - yet from behind a "veil of ignorance" about their own personal situation (Rawls 1971).

The most popular variant of the impersonal social decision-maker approach is known as the "Person-Trade-Off" (PTO) approach (Nord 1995). Respondents are asked to choose between two groups of people who differ in certain specified respects (e.g. age, lifestyle, health) and who

---

29 These equity weights are not to be confused with the further age-weights that Murray and Lopez apply to account for the value of people's differing social roles (e.g. being a parent) at different stages in their lives, which is a utilitarian efficiency consideration rather than an equity consideration.
stand to benefit from some pair of alternative health interventions. The trade-off is established by asking them to vary the number of people in one group while holding constant the number in the other group, so as to find a point of indifference. Valuations from this technique can be seen as representing the trade-offs that people are prepared to make between health gains for different kinds of people and numbers of persons treated. A less popular variant of this approach is to vary the health gain (or distribution of gains) for one group while holding everything else constant (including the number of people in each group).

Whichever variant is used, the great difficulty lies in defining the value elicitation questions tightly enough to elicit precisely targeted information on equity-efficiency trade-offs of a kind which can be used in economic analysis. As we shall see, a number of studies have obtained results which cannot be interpreted as equity-efficiency trade-offs, since they fail to distinguish between equity and efficiency considerations sufficiently clearly. Furthermore, they may also fail to distinguish different equity considerations. Responses to PTO questions, for instance, are likely to contain the relative weights a respondent attaches to a number of potentially relevant equity factors (in particular, the severity of the pre- and post-intervention health states, the health gain as a result of intervening, and the number of persons treated) and it may be impossible to disentangle what are the relative weights attached to each of these considerations. The weights attributed to these different equity considerations may have quite different implications for resource allocation decisions.

This problem of confounding factors is compounded by the psychological difficulties faced by respondents who are not used to thinking about the highly focused and simplified questions about equity and fairness that economists tend to ask. Consequently, they may well interpret questions in unanticipated ways, and bring to bear wider considerations that the economist wishes to set aside (Dolan and Cookson 1998). In such circumstances, it is often hard to know whether or not respondents would accept the real-world policy implications of their hypothetical answers, once those answers have been fed into the economist’s theoretical model and translated into practical policy guidance. Thus questionnaire studies need to steer a somewhat heroic course between the Scylla of confounding factors and the Charybdis of befuddled respondents. Given these twin perils, the handful of published numerical trade-off results summarised below must be viewed with much caution.

The issue of age is the only area in which economists have come close to estimating a practically useful equity-efficiency trade-off parameter. Studies have consistently found that people are, broadly speaking, willing to discriminate in favour of the young and against the elderly in distributing health care treatments (Charny, Lewis and Farrow 1989, Nord et al. 1996). Two studies have estimated the size of this trade-off, using large-sample telephone

---

30 This requirement means that questionnaires used by economists tend to be far more abstract and tightly defined than those used by opinion pollsters and sociologists.

31 Dolan (1998) suggests an alternative approach which uses a particular class of health-related social welfare functions (HRSWF) which allows efficiency and equity to be considered independently. The function has two parameters: one which determines the general degree of aversion to inequality (the curvature), and one which determines the specific equity weight given to one class of recipients over another (the slope). Unlike the PTO methodology, it allows us to address two key issues separately: 1) how individuals value one health state compared with another, and 2) how society values those same health states. Dolan argues that it is important that these two issues are kept separate since each may need to be answered in a different way in different contexts.

32 More specifically, Charny et al. (1989) find an "inverted-U" shaped valuation of life with age: the value rises from a 2-year old up to an 8-year old (on the grounds that the 8-year old is "more of a person"), but then falls back down again from a 35-year old to a 60-year old (on the grounds...
surveys, by asking people to choose between two life saving programmes which save different numbers of people of different ages. Cropper et al. (1994) find that their median respondent was willing to let seven 60-year olds die to save one 20-year old; Johanneson and Johannsson (1997) find that their median respondent was willing to let forty-one 70-year olds die to save one 30-year old.

Unfortunately, these estimates cannot be interpreted as equity-efficiency trade-offs as they stand, because they fail to distinguish between efficiency reasons for ageism (e.g. the utilitarian consideration that young people typically gain more health from treatment than elderly people) and equity reasons (e.g. the "fair innings" argument). So Johanneson and Gerdtham attempt to infer the relative strength of equity and efficiency considerations through empirical modelling techniques. By making the heroic assumption that respondents based their answers upon average quality-adjusted life expectancy for people of that age, they infer that nine QALYs gained for 70-year olds are judged equivalent to one QALY gained for 30-year olds.

The only other equity trade-off estimate we are aware of was made by Johannesson and Gerdtham (1996), who piloted a version of the "veil of ignorance" approach. Using a telephone survey, they asked respondents to choose between two societies that differ with respect to the number of QALYs received by two groups. They found that, on average, respondents were willing to give up 1 QALY in the group with more QALYs to gain 0.45 QALYs in the group with fewer QALYs (irrespective of how many fewer QALYs). This result is hard to interpret, however, in the absence of any supporting qualitative evidence about what background assumptions respondents were making about other equity considerations (especially the age of people in the different groups).

Although no trade-off estimates have been made, economists have also begun to investigate the issue of how far respondents are willing to discriminate against those who are partly responsible for their own ill-health. So far, it seems that the majority of people do wish to discriminate against those with purportedly "self-inflicted" illnesses, both in group choices and in clinical decisions (Kneeshaw 1997, Dolan and Cookson 1998). However, (i) a substantial minority dispute this principle, (ii) there is more agreement on the general principle than on its specific application to groups such as smokers, heavy drinkers, the overweight and illegal drug users, and (iii) following group discussion, people are less likely to support this principle when they become more aware of the conceptual and practical difficulties associated with it (Dolan and Cookson 1998).

In conclusion, it is fair to say that this empirical work is still at an embryonic stage, and that no-one has yet devised a fully satisfactory questionnaire method for eliciting equity-efficiency trade-offs in health. It is worth noting that this is by no means a unique failing of

---

33 Two issues of procedural equity (which are outside our remit) have also received particular attention: (i) equality of access to health care and (ii) urgency of treatment. It has been found that people give substantial weight to equality of access and are not willing straightforwardly to give higher priority to some patient groups purely on the basis that they stand to gain more health from treatment (Nord 1993, Nord and Richardson 1995, Dolan and Cookson 1998). It has also been found that people wish to give higher priority to those in life-threatening emergency situations, independently of capacity to benefit, both in rankings of specific health care services (Kneeshaw 1997) and in clinical choices (Dolan and Cookson 1998).

that the 35-year old is more likely to have parental responsibilities). Since this study used scenarios involving clinical choices between two patients, however, it is not clear how far this finding is transferable to the settings we are interested in, involving choices between groups made by managers and policy makers.
health economists. Economists have examined the distribution of a number of things other than health - from income right through to grapefruits - without managing to elicit reliable equity-efficiency trade-offs (reviewed in Miller 1992). The basic finding from this work is that people do not subscribe to any one "pure" theory of distributive justice, applicable in all circumstances. Rather, people's responses appear to mix different concepts of equity together (e.g. some weight on procedural justice, some on attaining better outcomes, some on meeting minimum constraints), and to vary the chosen equity concept from one context to another.  

The great challenge is to bridge the gap between the economic requirement to estimate precisely targeted equity-efficiency trade-offs, and the psychological capabilities of respondents to think about equity and efficiency in such a tightly defined manner. In part, this will require greater sensitivity of economic theory to the philosophical subtleties and context-sensitivities embedded in people's preferences. It will also require more sophisticated methods for eliciting preferences which are capable of teasing out key considerations from confounding factors. Researchers have begun to respond to this latter challenge by combining questionnaires with in-depth group discussions, which (a) allows respondents more time to digest and to reflect upon the questions, and (b) gathers qualitative evidence about how respondents have interpreted the questions (Nord et al. 1996, Dolan and Cookson 1998).

VIII FURTHER CONSIDERATIONS

At the level of principle the main weakness of the work reported here is that it regards health as the sole element in the social welfare function, and ignores other objectives of public policy. This is defensible to the extent that public policy responsibilities do in reality tend to get fragmented, and in this case a concern for health is at least much wider than a concern for health care. It might also be defended on the ground that if it were only possible to work with one indicator of people's well-being, people's lifetime experience of health, measured in QALYs, might be a better measure than any other (see Sen 1998). A person's quality-adjusted-life expectancy is more individualised than (household) income or wealth, and is known to reflect differences in people's income, education, occupation, location and general economic and social environment. Health is an essential ingredient for every individual's well-being, no matter what values and aspirations they have. So treating it as the primary concern of public policy may be quite sensible and revealing, even though somewhat unorthodox.

The weakness of so doing, of course, is that it denies us the opportunity to explore the extent to which people will trade off health against other desirable things - such as wealth, leisure, pleasure, social status, self-respect, friendship, and so on. If, for instance, individual well-being is seen as holding a balanced portfolio of health, wealth and wisdom (see Williams 1988) then all three need to be in both individual and social welfare functions. Indeed, the United Nations (UN Development Programme 1992) has in fact devised an index for the comparison of levels of well-being between nations that combines life expectancy, real income and literacy levels, with that specific purpose in mind. But from an equity viewpoint it may well be that people have different, rather specific, views about equity in each of these realms. For instance, they may be very egalitarian concerning health, but not at all so with respect to income and wealth. So it cannot be assumed that an equity notion which is strongly supported (and well documented quantitatively) in the health field should also be applied to these other fields .... and

---

34 One particularly important aspect of the context is that preferences depend upon the status quo distribution of outcomes (Kahneman, Knetsch and Thaler 1986, Bukszar and Knetsch 1997). That is, outcomes which impose losses of income on others relative to a status quo situation are perceived as being unfair, quite apart from egalitarian considerations relative to absolute levels of income.  

35 A view elaborated by the political philosopher Walzer (1983).
vice-versa!

At a more practical level, economists need to widen their repertoire of equity notions in the health field, and be rather more careful about specifying why a particular aspect of health (e.g. someone's current health state) should be of policy interest from the viewpoint of distributive justice, when presenting data (usually cross-sectional) about inequalities in health. It would be helpful if it became the norm to link any statement about equity to some specified philosophical position, so that the reader could see more clearly how the data should be interpreted, and what are the strengths and weaknesses of the author's position at a conceptual level.

In a day to day setting, there is likely to emerge some division of labour between economists according to their immediate responsibilities. Many health economists are engaged in economic evaluations of health care activities, and employ methods of analysis that are designed to elucidate the relative efficiency of different options (in cost-effectiveness, cost-utility, or cost-benefit terms). They can hardly be expected, in the context of each such study, also to embark upon the generation of all sorts of different equity weights for the diverse clientele of public and private agencies who might one day want to use their results. The most that can reasonably be expected of them is that they make it possible to identify the distribution of health, and of health benefits, across the subjects in their study, by the main personal characteristics that are likely to be relevant to judgements about distributive justice. It is likely that this will need to go beyond the usual background data on age and sex and area of residence, and include factors such as lifestyle characteristics, occupation, education, income, ethnic group etc.

Meanwhile a second group of economists need to be working on the development of the equity concepts themselves, and on estimating the shadow prices and equity-efficiency trade-offs they imply or require. The objective should be to derive context-specific equity weights that could then be applied by policy analysts to the data generated by the efficiency-based evaluative studies, to ensure that considerations of equity in health do not get lost, but get treated in a systematic (rather than an arbitrary and capricious) way, when the results of an efficiency calculus are up for consideration. By harnessing methods from psychology, sociology and philosophy to this process, health economists may in the future succeed where others have failed in measuring equity-efficiency trade-offs. It will be a major step forward to find a way of translating people's considered preferences about equity in health into measurements that can give clear and concrete guidance to social decision-makers in this field.

Acknowledgements

This chapter has benefited from the comments of various colleagues, to whom we are most grateful for the time they have spent reading and thinking about what we had written at draft stage. For this help we would like to thank the following: Gwyn Bevan, James Buchanan, Tony Culyer, Diane Dawson, Paul Dolan, Lucy Gilson, Julian LeGrand, Susan Mendus, Richard Musgrave, Peter Singer, Michael Spackman, and Aki Tsuchiya. A special additional word of thanks goes to Vanda Castle, who coped stoically with the frequent changes we made to her painstakingly constructed diagrams.
References


Hare, R.(1993),"Health care policy: some options” in Hare, R., Essays on Bioethics , Calendon, Oxford, p 210


Smart, J.J.C. and Williams, B. (1973), Utilitarianism: For and Against (Cambridge University Press).


