

Why Not Just Ask? Preferences, “Empirical Ethics” and the Role of Ethical Reflection

Daniel M. Hausman

University of Wisconsin-Madison

Draft of October 12, 2000; forthcoming in a WHO volume edited by C. Murray and D. Wikler

Many questions concerning health involve values. How well is a health system performing? How should resources be allocated between the health system and other uses or among competing health-related uses? How should the costs of health services be distributed among members of a population? Who among those in need of transplants should receive scarce organs? What is the best way to treat particular patients? Although many kinds of expertise bear on these questions, values play a large role in answering them. These values are of many kinds – judgments about how health states contribute to the well-being of individuals depend on views about what is good for a person, whereas judgments about how to share the costs of health care or about how to distribute scarce organs among those in need of transplants depend on theories of fairness. The value judgments that need to be made are not only multifaceted and heterogeneous; they are also controversial. Both the understanding of the questions and the substance of the answers may vary widely from individual to individual and among different groups within countries and across national boundaries.

How are these many complicated, difficult and controversial evaluative questions to be answered? This could be a question about *who* should answer them as well as a question about *how* they should be answered. There are various possible mechanisms, and there is no reason why the same method should be employed to answer every question. Among the possible methods are (1) moral argument, (2) political deliberation and (3) eliciting preferences or judgments from some group. Included in the third method lie many different techniques of elicitation. For example, one could study attitudes by the use of deliberative (focus) groups, by permitting individual choice in some sort of market or quasi-market, by polling or by voting. These methods of answering evaluative questions are not always incompatible with one another. Moral argument can play a role *within* political deliberation, deliberation by focus groups, individual market choice, or voting. Polling can influence political deliberation, and so forth. Furthermore, these methods of resolving evaluative questions concerning health care are not well defined until one specifies *who* will be polled, who will argue, who will vote, or who will choose. Polling a random sample of the whole population to determine the seriousness of a disability is a very different method than polling a sample of health-care professionals and is unlikely to give the same answer.

In the face of all this complexity, which is only magnified when one recognizes the variation

across cultures, it might appear that the only sensible alternative is to rely upon the values that prevail within each culture. These may be difficult to elicit, but the difficulties are sociological and psychometric, not ethical. The task is not to figure out how the cost of health care *ought to be* distributed or whether for example blindness *is* a more serious disability than deafness (in a particular society). Instead the task is to figure out what is the social consensus on these questions. The fact that the answers to evaluative questions about health and health care have to be accepted within individual cultures makes the replacement of ethical questions by sociological questions about prevailing attitudes all the more tempting. Indeed those concerned with health policy may even forget that there are real ethical questions here, and that the answers that are widely accepted within any particular culture are not guaranteed to be correct. Thus, for example, in Erik Nord's recent book, questions about how to allocate health care are assimilated to questions about how best to serve the objectives of those whom the health-care system serves. He writes, for example, "I define a *fair* resource allocation in health care as one that accords with societal feelings about the strength of claims of different patient groups" (1999, p. 23).

Although I just quoted Nord's recent book and will refer to it again later, this essay is not meant to be a critique of Nord. On the contrary, despite the statement just quoted, Nord is acutely aware of the need for ethical reflection;¹ and he would not maintain that ethical questions about health and health systems can be replaced with sociological questions concerning prevailing attitudes. As a matter of fact, I do not know of anyone who has argued in print that there is *no* role for ethical reflection in addressing the myriad problems concerning health and health systems. But the line of thought that these difficulties can be solved if the World Health Organization will "just ask" is a tempting one and deserves a full and careful answer. Why shouldn't those concerned with health policy simply ask those who are affected by the policy? Why aren't the difficulties in answering evaluative questions merely technical problems about how to elicit the responses of the target group? What role should ethical reflection play, and what part should be taken by those who have devoted themselves to ethical reflection, such as religious leaders, academics, essayists, journalists, and even some politicians?

These questions about how to understand and to answer evaluative questions about health and health policy are very complicated, and I shall have to touch on a wide variety of issues ranging from the nature of morality to the methods of economics. In order to narrow the topic, I shall focus on questions concerning the evaluation of states of health and of health improvements. The reason for

¹ See for example Nord 1999, p. 90 and Menzel, Gold, Richardson, Nord and Pinto-Prades 1999. The latter emphasizes the role of bioethicists and focuses on ethical factors that have been left out of cost-effectiveness analysis.

concentrating on these questions is that the case for “just asking” seems strongest here. It seems obvious that people’s preferences for different health states should have at least *something* to do with their value. Nevertheless, I shall argue that ethical reflection has an essential role to play even with respect to the evaluation of health states. Most of what I shall say about the need for ethical reflection applies unproblematically to more obviously ethical questions about the allocation of health care resources or the distribution of costs, and later I shall say a few words about the application of my arguments to more questions such as these.

In the first three sections of the paper I shall suppose that the valuing of health and health care depends entirely on the preferences of those affected by health policies. Later I shall argue against this view, but first I will examine the task of answering evaluative questions concerning health and health policy from this perspective. Section 1 explains how most economists conceive of the relationship between preferences and well-being and how this view leads them to interpret comparisons of health states as matters of preference. Section 2 presents the specific theory of preference-based valuation defended by most economists and explains why health economists have taken a somewhat different approach. Section 3 discusses some actual methods of eliciting values for health states, and it argues that philosophical reflection has a role, even if one supposes that the task is merely to determine what are the preferences of the target population. Section 4 then distinguishes between preferences for health states and judgments about the relative severity of health states and argues that when health economists have claimed to be measuring preferences, they have in fact often been eliciting opinions. Section 5 explains why both preferences and opinions are nevertheless relevant to the evaluation of health states. Section 6 argues that the evaluation of health states and interventions is not settled by the judgments of those who are concerned. Section 7 addresses arguments in defense of relying on the values of the target population which do not rely on the relativist view that socially accepted judgments are automatically correct, but which instead emphasize the importance of popular acceptability or sovereignty. Section 8 draws together my objections to answering evaluative questions by simply asking people what they prefer or how they would judge and clarifies the respective roles of philosophical reflection, the elicitation of popular judgments, and the measurement of preferences.

1 Preferences, values and well-being

One crucial element in the WHO’s study of health and health systems is the effort to specify a summary measure of population health, which will allow comparison of health in different countries and in different groups and regions within countries. In addition, such a summary measure of

population health could contribute to the assessment of interventions to improve health, since health improvements are, of course, differences in health levels. Developing a summary measure of population health requires some way to assign numbers to different illnesses or disabilities so that one can add up the burdens of poor health and mortality and thereby measure overall health.

At first glance, it might not be obvious that this effort to render health states comparable and then to measure health involves values at all. "Health" seems a medical or biological notion, linked perhaps to some notion of normal functioning (Boorse 1975, 1977; Daniels 1985, pp. 28-30). On such a conception of health, the comparison of health states would appear to be a purely technical exercise devoted to determining the extent to which different health states diminish the range of functioning.

Even if this conception of health were correct, there is no way to weight the different domains of functioning without evaluative commitments concerning what constitutes a good life. The notion of health cannot be defined or measured independently of some notion of what constitutes a good life. Physical and mental states that diminish one's prospects of having a good life are states of bad health, while physical and mental states that improve one's prospects are states of good health. Both the question of what a good life is – of what well-being consists in – and questions about how much different health states diminish the goodness of life are evaluative matters.

Health and well-being are of course not the same thing: *A* may be healthier than *B*, even though *B* is better off than *A*. But better health correlates strongly with greater well-being, because health is an important component of well-being, because health causally contributes to other components of well-being, and because other components of well-being contribute to health. When people compare two very different health states – for example, when they compare the state of being depressed to the state of having recurrent gastro-intestinal symptoms – they must, at least in part, be comparing how well or badly off someone is when depressed relative to how well or badly off one would be with diarrhea and stomach cramps.

The crucial point is that a person's state of health cannot be evaluated in purely biological terms. It is not just a matter of biological functioning or evolutionary fitness, because many important health problems have comparatively little to do with functioning and virtually nothing to do with fitness. And even if health were entirely a matter of functioning, one would have to evaluate the relative significance of limitations in different domains of functioning. Furthermore, since human traits are functional only if they are good for people, there is no way even to specify the functions of human traits without reference to what makes for a good life. Is someone with an acute sense of smell in better or worse health than someone without one? Even after one specifies the environment, it may not be obvious whether the consequences of an acute sense of smell are

“beneficial” and thus possibly *functions* of an acute sense of smell. To decide whether the consequences are beneficial requires a theory of well-being.

Moreover, defining some measure of population health requires more than merely being able to say when well-being increases or decreases. When considering whether *A* is healthier than *B* at some particular instant, one does not need to ask “how much healthier.” But if one wants to compare the health of whole populations at an instant, then (unless everyone in population I is no less healthy than anyone in population II), one needs to consider the distribution of different states of health in the two populations and hence how *much* healthier some individuals are than others.² In addition, even if one is comparing only two individuals and is able to say whether *A* is healthier than *B* or vice versa at every instant during some time interval, one would not necessarily know whether on average over the interval *A* is healthier than *B*. To compare *A*’s health to *B*’s over some period during which *A* was sometimes healthier than *B* and *B* was sometimes healthier than *A*, one needs to know how *much* healthier *A* was when *A* was healthier and how much healthier *B* was when *B* was healthier. It is also necessary to take into account the amount of *time* that *A* was healthier than *B*. The obvious thought is that one should graph *A*’s and *B*’s health – or the difference between their health – on the vertical axis and time on the horizontal axis and derive the comparison between their health over the whole period by integrating. Even if we will never be able to carry out this procedure with precision, perhaps some reasonable approximation is possible.

Recognizing the connections between health and well-being or quality of life only helps one to measure health states if one has a theory of what contributes to a good life. Exactly what is good for a particular person, *A*, will depend on *A*’s character, ability, and circumstances, and what is good for *A* may be very different from what is good for *B*. But most of the differences between what is good for *A* and *B* concern *instrumental* goods, things that are good because they are *means* to something else. If one focuses on *intrinsic* goods, things that are good in themselves, without regard to their consequences, then there may be much less variation from individual to individual. Notice that there must be intrinsic goods in order for there to be instrumental goods. One central question of moral philosophy has been to determine what things are intrinsically good for human beings. Aristotle, for example, held that happiness was the sole intrinsic good (1962, book I). Since health seems both an intrinsic good and an almost universal means toward other intrinsic goods, one should expect a good deal of common ground among people concerning the value of health.

In chatting with one’s neighbors, as in studying moral philosophy, one finds many different

² There is one possible exception. If one were to assess the health of a population by comparing the health of those who are healthiest (or least healthy) in the two populations, then there would be no need to compare differences in health states. But these are not the methods of comparing population health that are at issue.

theories of well-being. In some religious views, the good lies in a relationship with God, while in others a relationship with God is good because of the eternal happiness it brings. Many people believe that only mental states are intrinsically good, but there is less agreement here than it seems, because there are so many different views of which mental states are intrinsically good. Jeremy Bentham holds that the good is pleasure (1988), while John Stuart Mill (1863) holds that it is a diverse set of mental states he calls “happiness.” Mystics find the good in contemplative states of mind. Henry Sidgwick argued for the hybrid view that the good is any mental state that is intrinsically desirable (1901).

There are also many defenders of non-mental-state views. Friedrich Nietzsche regards great achievements as the ultimate goods. T.H. Greene finds the ultimate good in self-realization (1969) . Others endorse as intrinsic goods a whole pot-pourri ranging from health and intimate personal relationships to achievements such as those Nietzsche admired (Griffin 1986). The theory of well-being is an unsettled area of philosophy. It is difficult even to categorize the various theories, and each faces serious difficulties. All of this is enough to frighten away non-philosophers. But philosophical problems concerning the constituents of a good life cannot be avoided, if one wants to be able to compare health states.

Theories of well-being can be classified as either “formal” or “substantive.” A substantive theory of well-being says what things are intrinsically good for people. “Hedonism” is an example of a substantive theory of well-being. It says that well-being is pleasure. “Formal” theories of well-being specify how one finds out what things are intrinsically good for people, but they do not say what those things are. To maintain that welfare is the satisfaction of preferences, as many economists do, is to offer a formal theory of well-being. This theory does not say what things are good for individuals, but it says how to find out – by seeing what they prefer. Formal theories may be compatible with substantive theories. For example, if happiness is what everyone most prefers, then the formal claim that well-being is the satisfaction of preference and the substantive thesis that well-being is happiness could both be true.

One might question whether there is anything to be gained from thinking about theories of well being, since there seems to be more consensus in untutored judgments about health than in philosophical accounts of well-being. Indeed, economists are famous for maintaining that interpersonal comparisons of welfare are impossible (Robbins 1935, chapter 6), but nobody asserts that interpersonal health comparisons are always impossible. Given the obvious and serious difficulties in the way of providing a theoretical method for valuing health states, it is tempting to surrender the attempt and instead simply ask the members of the group whose health is in question to value the health states themselves. I shall argue that this proposal is a subjectivist counsel of despair.

It is subjectivist and, as we shall see, relativist, because one is abandoning the effort to determine what the values of different health states truly are and substituting an inquiry into how those affected value health states. This replacement of a theory of valuation by measurement of attitudes of the target population is a counsel of despair, because if theorists cannot figure out any justifiable way of comparing health states, there is little reason to suppose that non-theorists will do better.

At this point, however, a crucial twist is introduced. Economists are inclined to interpret evaluative questions such as, “Is it worse to be blind or deaf?” or “Does it contribute more to population health to cure or prevent X cases of blindness or Y cases of deafness?” as questions about preferences such as, “Do people prefer to be blind or deaf?” and “Are preferences better satisfied when X cases of blindness or Y cases of deafness are cured or prevented?” If evaluative questions were questions concerning preferences, then it would be perfectly reasonable to answer them by measuring preferences.

If one believes, as most mainstream welfare economists do, that value is constituted by preference, then questions about the value of health states *are* questions about preferences. There is a fact of the matter about how good or bad it is to be blind or deaf, and this fact is determined by people’s preferences. According to this view, it is worse for me to be blind than to be deaf if and only if I prefer being deaf to being blind. Just as we can investigate whether people have higher blood pressure when they eat salty foods than when they do not by measuring people’s blood pressure, so we can determine whether people have a higher level of well-being when they are blind than when they are deaf by measuring their preferences.³

Although some economists would like to think that by taking well-being to be the satisfaction of preferences they can avoid philosophical commitments concerning the theory of value, in fact they are adopting a specific philosophical theory – a formal theory of the good as the satisfaction of preferences. Although this is undeniably a theory of well-being, it seems to many economists a great deal humbler than substantive theories, which attempt to say what things are intrinsically good or bad. By leaving the substantive question of what is good for an individual up to the individual, economists and health professionals believe they are showing philosophical modesty. The preference-satisfaction view of well-being also appeals to the anti-paternalist values of many economists. Nevertheless it is questionable whether the preference-satisfaction theory is less philosophically controversial than substantive theories.

The preference-satisfaction theory of well-being is connected to the way in which most

³ There are also philosophical reasons for associating values and preferences that derive from a subjective view of value (Gauthier 1986, ch. 2). In any plausible subject philosophical theory of value, there is, however only a distant and very complicated relationship between values and preferences.

economists explain and predict choices. In standard economic models, individuals are supposed to have complete and transitive preference rankings. What this means is first of all that given any two alternative objects of choice x and y , individuals can say whether they prefer x to y , y to x or whether they are indifferent. Second, their preferences are transitive: if they prefer x to y and y to z , then they prefer x to z . So individuals can in principle consistently list all the objects of choice, with x in a higher row than y in the list when x is preferred to y and x and y in the same row when the individual is indifferent between them. One can then assign numbers to the rows, with higher rows getting higher numbers. These numbers – whose only significance is to indicate position in such a list – are what economists call “utilities.” “Utility” was for a long time used to refer to the properties of objects that cause desirable mental states or to those mental states themselves, but in contemporary economics, utility is just a way of referring to preference. It is nothing substantial and certainly not itself an object of preferences. Maximizing utility is just doing what one prefers.

In addition to maintaining that individuals have complete and transitive preferences, the standard model takes an individual's choices to be determined by the individual's preferences. This sounds as if it involves self-interest, but it does not. If I prefer to sacrifice my interests to yours and act on that preference, then I am acting according to my preferences, but against my self-interest. Economists do, of course, usually assume that individuals are self-interested, but this is not built into their basic model of rational choice.

The standard theory of choice says nothing about well-being or about ethics, but it is surprisingly easy to get from it to the normative view that well-being is the satisfaction of preferences. Assume that individuals are self-interested (as economists often do) or that what is at issue mainly involves oneself (as is often the case in comparing health states). If individuals are exclusively self-interested, then they will prefer x to y if and only if they believe that x is better for them than y is. If they are well-informed, then their beliefs will be true, and they will prefer x to y if and only if x is better for them than y . So it is very tempting to take well-being to be the satisfaction of preferences.

There are however many objections to a preference-satisfaction view of well-being. Real individuals are not exclusively self-interested. People are sometimes altruistic and all too often malevolent. Real individuals are also ignorant of many things. So people may prefer something that is bad for them because they mistakenly believe it is beneficial or because they want to help a friend or harm an enemy. It is not true that x is better for A than y if and only if A prefers x to y . These objections are not hard to see. It takes no great philosophical talent to recognize that giving an automatic weapon to a reckless teenage boy does not necessarily make him better off, regardless of how much he wants it.

Yet welfare economists continue to espouse the preference-satisfaction theory of welfare. Why?

One reason is the argument just given: if one accepts the standard view of rationality and the standard idealizations of positive economics – self-interest and perfect knowledge, then it follows that what one prefers is what is good for one. So within the theoretical world depicted in many economic models, welfare is preference satisfaction. Economists recognize that this world is not the real world, and the fact that welfare is preference satisfaction in standard models does not imply that welfare is preference satisfaction in real life, but they often regard the differences between theory and reality as matters of detail. “If one looks past the complications of actual life to the central realities captured in standard economic models, one can see that welfare is in essence the satisfaction of preferences.” This line of thought will not convince anyone who does not see the world the way mainstream economists do, but it helps explain why these economists are content to identify well-being with the satisfaction of preferences.

A more relevant reason is that in the context of comparing health states, the interests of others may be relatively less important, and specific tactics can be employed to limit errors. For example, after the unfortunate attempt in Oregon to elicit health preferences by telephone surveys, most health economists would now insist that one has to provide people with a good deal of knowledge and considerable opportunities to reflect and to recognize their mistakes before one can successfully elicit their “true” preferences. Perhaps one can in this way render errors relatively unimportant.

In addition, even though what satisfies A 's preferences does not necessarily make A better off, A 's preferences may be the best *guide* to what is beneficial to her. Even if satisfying people's preferences does not always make them better off, there is perhaps no better way to decide how to benefit them. Those designing and implementing health policy know less of A 's circumstances than she knows and they typically have a less fervent concern for her well-being than she does. Despite their technical expertise, which may make them a better judge of matters such as which medication A should take, the judgments of health-system administrators about how to make A better off are often likely to be worse than her own judgments. Furthermore, even if A 's preferences are not a better guide to what is good for her than the judgments of administrators, one might argue that it is *safer* to leave the judgment to her. This defense does not attempt to show that welfare *is* the satisfaction of preferences. Instead it denies that economists need any philosophical theory of welfare. Regardless of what human well-being truly is, the best *measure* of well-being is the extent to which preferences are satisfied. But there is of course no way to defend the claim that preference satisfaction is the best measure of welfare if one has no idea what welfare is. So the need for ethical reflection remains.

I would guess that in fact many economists do not take the standard definition of welfare literally. They believe instead that welfare is a desirable mental state, such as happiness. They find it plausible to believe that the best measure of well-being is preference satisfaction, because they believe

that the best way to implement a social policy that aims to make people happy is to satisfy preferences. I cannot prove this conjecture that most economists actually take welfare to be a mental state like happiness, but there is a good deal of evidence that supports it. Economists often slide from talking about utility to talking about happiness. They often talk about individuals “seeking” utility, which makes no sense if utility is just a measure of the extent to which preferences are satisfied. And it is easy to equivocate on the word “satisfaction.” To satisfy a preference is for a preference to come true. It has nothing to do with any *feelings* of satisfaction. *A*’s grandmother’s preference that her granddaughter become a doctor is satisfied if *A* becomes a doctor, even if *A*’s grandmother never lives to see that day and cannot *feel* any satisfaction at the event. But economists often speak of satisfaction as a feeling.⁴

Even if the theoretical and practical defenses of measuring welfare by the satisfaction of preferences do not succeed, they permit one to see why sensible economists so readily link welfare and the satisfaction of preferences. Let us now look more carefully at the consequences of a preference-satisfaction view of welfare for economic evaluation.

2 Economic valuation

When economists attempt to evaluate economic institutions, policies and outcomes, they usually ask whether they make people better off. Given their theory of welfare as the satisfaction of preference, they are asking how well institutions, policies and outcomes satisfy preferences. However most economists deny that objective comparisons can be made of how well satisfied are the preferences of different individuals. So they cannot assess policies by adding up their consequences for the welfare (that is, preference satisfaction) of different individuals. Indeed it seems that alternative policies and institutions can only be compared when one of them satisfies at least one person’s preferences better without satisfying anyone else’s preferences less well.

It is for this reason that *efficiency* has such a special importance in standard welfare economics. “Efficiency” in theoretical economics is not fundamentally a matter of saving time or materials in production or distribution (LeGrand 1991). An efficient state of affairs (which is often called a “Pareto optimum”) is one in which it is impossible to make someone better off (in terms of preference satisfaction) without making someone else worse off. The character of welfare economics is determined by its focus on welfare, its identification of welfare with the satisfaction of preferences,

⁴ The view that well-being is a mental state such as happiness is however just about as implausible as the view that well-being is the satisfaction of preferences. Robert Nozick argues forcefully that we should not exchange the possibility of having real contact with the world in order to have the highest quality mental states (1974, p. 41).

and its rejection of interpersonal welfare comparisons.

If people are better off being blind rather than deaf if and only if they prefer being blind to being deaf, then measuring preferences is a way of determining which of these health states is truly worse. This reason for "just asking" is radically different from the view that evaluative questions are really sociological questions and thus to be answered by sociological inquiry. The economist's view relies on a significant normative premise: that well-being is the satisfaction of preferences. So in determining whether people would prefer to be blind or to be deaf, one is determining whether those who are blind are worse off than those who are deaf. One is not eliciting opinions concerning the goodness of these two health states and then concluding that the consensus judgment is automatically correct. One is instead eliciting *preferences*, and because preference satisfaction constitutes well-being, the preferences *determine* which health state is *in fact* better – regardless of anybody's opinion on the matter.

If evaluative questions concerning health and health policy are questions concerning preference, then it might appear that they can be answered by asking people what their preferences are. But almost all economists would deny that naively asking, "Would you prefer to be blind or deaf?" would enable one to value these two health states. There are four main reasons why. First, not everyone will answer the same way, and nothing has yet been said about how to aggregate preferences. Even if it follows from a respondent's preference to be blind rather than deaf that it would be better for that individual to be blind than to be deaf, unless everyone's preferences agree, one has no basis for an overall or average ranking of the health states themselves. Nor can one simply count heads and take the majority preference as decisive, because mere expressions of preference between health states provide only an *ordering* of those health states. What if the preferences of the minority who prefer being blind to being deaf are *stronger* than the opposing preferences of the majority? Are preferences on average better satisfied by satisfying a greater number of weaker preferences or by satisfying a smaller number of stronger preferences? For purposes such as generating a summary measure of population health, one needs to know how *much* better or worse one health state is than another.

These two problems lead to a third. If the overall value of health states such as blindness and deafness depends both on how *many* prefer one to the other and on how *strong* these preferences are, then it will be necessary to make interpersonal comparisons between the strength of the preferences of different individuals. Nothing has yet been said about how to make interpersonal comparisons, and many economists deny that interpersonal comparisons of well-being or preferences are possible. Fourth, economists are skeptical about individual expressions of preferences and consequently concerning survey results. Indeed economists who accept revealed-preference theory

(Sen 1971, 1973; Hausman 2000) are uncomfortable with a subjective notion of preference and would like to be able to define preference in terms of choice. Economists are accustomed to paying special attention to market data, and they are cynical about expressions of preference ("mere words") unless they influence behavior.

According to the vision that animates mainstream economics, when there are markets in which everything one is valuing can be bought and sold, then these problems are solved automatically. Those who most strongly want a commodity will be willing to pay the most for it. Those who want the commodity least will accept the lowest price to part with it. Eventually a price will be reached so that everybody who wants the commodity at that price will be able to purchase it, and everybody who prefers to part with the commodity at that price will be able to part with it. Such a market-clearing price responds both to the numbers of people who want to buy or sell and to the intensity of their preferences. It does so by measuring the intensity of everyone's preference for one unit more or less of the given commodity by the amount of money people are willing to pay for it.

This measure of intensity of preference and this implicit interpersonal comparison of preferences is imperfect, because people do not value money equally. There will be systematic errors, because a few dollars matter a great deal more to somebody who has very little money than to somebody who is wealthy. The price of a good that is wanted mainly by the poor will understate the good's social value.⁵ Comparison of the prices of two commodities permits a comparison of the social values of one unit more or less, but only *given* the existing distribution of wealth. Economists grant this consequence, but many would say that when market values are askew, the problem is with the distribution of wealth, not with the market as mechanism for valuing things.

Notice that only the social value "at the margin" – that is the social value of consuming one unit more or less is defined. But since choices can often be decomposed into choices of consuming one more or one less unit, information about value at the margin often suffices. Even though value at the margin is a measure of preferences, it also depends on other things, such as wealth and costs of production. For example, as clean water becomes more expensive to procure, the value of water goes up. It goes up not because people's underlying preferences change, but because when water is more expensive, people consume less of it. Because of this lowered consumption, an additional gallon of water is more valuable. Preference for a marginal unit is not a given, fixed independently of

⁵ An extreme example of this is the drug eflornithine, which is a highly effective "miracle" cure for sleeping sickness. Until 1999 the drug was produced by a U.S. subsidiary of Aventis, but when eflornithine proved ineffective against cancer (its intended target), Aventis stopped making the drug and gave the production license to the WHO. Only in early 2001, when stocks of the drug were almost exhausted, was the WHO able to find drug companies to manufacture it -- and then only because the companies hope to make profits from marketing eflornithine as a cream that removes facial hair. Because the victims of sleeping sickness are so poor, the amount

economic processes, but a consequence of underlying preferences and all the other determinants of prices and quantities supplied and demanded.

None of these claims about markets would appear to have much direct relevance to the valuation of health states or changes in health states, since comparatively few health improvements are purchased on the market by uninsured patients. What welfare economists propose when markets are missing – when for one reason or another things cannot be bought and sold – is that policy analysts simulate their operation. Economists need to infer willingness to pay indirectly from other market choices (which is much of what cost-benefit analysis consists in) or, as a last resort, to measure it via carefully designed surveys. When it is feasible to infer the complete demand and supply schedules for some commodity or service for which there is no market, economists can say what the market-clearing price and quantity would be.

Even in such a case, there is no way fully to mimic the operation of the market, since there is no feasible way to determine who would purchase the good and who would not. Markets for some commodity or service are usually missing precisely because the commodity or service cannot be privately or exclusively consumed, and the task for economists is consequently often to determine whether some commodity or service ought to be provided to everyone in some specified group. Thus, for example, economists could try to determine whether the costs of inoculating everyone against measles are larger or smaller than the economic benefits. The economic benefits could be calculated by examining losses of productivity, wages and so forth caused by measles or by attempting to measure directly what individuals would be willing to pay to be free of measles. Since it may not be feasible to provide the vaccine only to some people and not to all, economic appraisal of the vaccination program may depend on a comparison of total costs and benefits rather than on a comparison of the costs and benefits provided by the last few vaccinations.

This is an example of cost-benefit analysis. Projects are judged by comparing their monetary benefits and costs. Since markets are incomplete (or else there would be no need to simulate their results in cost-benefit analysis) one cannot simply read off the costs and benefit. Instead one needs to measure willingness to pay either directly via surveys or interviews or indirectly by attempting to estimate it from other market behavior.

Both to illustrate how economic valuation works and to show why it is unsuitable for many applications concerning health and health policy, consider a notorious example. In December of 1991, Lawrence Summers, who was then the chief economist at the World Bank, sent the following memorandum to some colleagues:

Just between you and me, shouldn't the World Bank be encouraging *more* migration of the

they are willing to pay for eflornithine understates its social value.

dirty industries to the LDC's [less developed countries]? I can think of three reasons:

(1) The measurement of the costs of health-impairing pollution depends on the foregone earnings from increased morbidity and mortality. From this point of view a given amount of health-impairing pollution should be done in the country with the lowest cost, which will be the country with the lowest wages. I think the economic logic behind dumping a load of toxic waste in the lowest-wage country is impeccable and we should face up to that.

(2) The costs of pollution are likely to be non-linear as the initial increments of pollution probably have very low cost. I've always thought that under-populated countries in Africa are vastly *under* polluted; their air quality is probably vastly inefficiently low [sic] compared to Los Angeles or Mexico City. Only the lamentable facts that so much pollution is generated by non-tradable industries (transport, electrical generation) and that the unit transport costs of solid waste are so high prevent world-welfare-enhancing trade in air pollution and waste.

(3) The demand for a clean environment for aesthetic and health reasons is likely to have very high income-elasticity. The concern over an agent that causes a one-in-a-million change in the odds of prostate cancer is obviously going to be much higher in a country where people survive to get prostate cancer than in a country where under-5 mortality is 200 per thousand. Also, much of the concern over industrial atmospheric discharge is about visibility-impairing particulates. These discharges may have very little direct health impact. Clearly trade in goods that embody aesthetic pollution concerns could be welfare-enhancing. While production is mobile the consumption of pretty air is a non-tradable.

The problem with the arguments against all of these proposals for more pollution in LDCs (intrinsic rights to certain goods, moral reasons, social concerns, lack of adequate markets, etc.) could be turned around and used more or less effectively against every Bank proposal for liberalisation (quoted in *The Economist*, February 8, 1992, p. 66).

Summers is not seriously proposing a World Bank program to export pollution to the LDC's. Such a policy faces technical difficulties, and there are moral arguments against it (though Summers points out provocatively that similar moral objections apply to other World Bank policies). This memorandum is of interest instead because Summers baldly put into words uncomfortable implications that most economists would prefer not to draw.

Air pollution and water pollution lessen health and the quality of life in many ways, yet most kinds of pollution – like most health states – have no market prices. Economists help guide choices in such cases by *imputing* costs. The hope is to figure out what costs *would be, if there were markets*

where pollution could be bought and sold. For example, economists may attempt to impute pollution costs by examining housing prices in communities that are much the same, apart from their air quality. They can draw inferences from how much people pay for air filters, water filters, or bottled water. They can collaborate with biologists in determining and assessing the costs of damage to health caused by pollutants. In such ways, economists may be able to estimate how much people in developed countries would be willing to pay to lessen pollution in their environment and how much people in LDC's would have to be compensated in order to be willing to accept more pollution.

People in a LDC might be willing to accept toxic wastes for very little compensation, because they were unaware of the contents of the wastes and the harms they might do. If their willingness to accept more pollution were based on such mistakes, it would not reflect what was really in their interest. But Summers does not rely on estimates of how eager people in developed countries are to be rid of their pollution or how willing people in LDC's are to accept it. What his three points attempt to show instead is that, insofar as they are *rational*, people in LDC's *should be* happy to sell pollution rights to people in developed countries for a price that the latter *should be* happy to pay. Summers argues that *the economic costs of the consequences of increased pollution are much lower in LDC's than they are in developed countries*. The willingness to accept more pollution in LDC's does not rest on mistakes about the consequences of doing so.

Suppose that environmental quality could be bought and sold in individual privately consumable units and consider whether rational and well-informed individuals, who live in a particular LDC, *Poor*, could strike deals to sell units of "environmental quality" to rational and well-informed individuals, who live in a developed country, *Rich*. If *Poor* is one of those "underpolluted" LDC's Summers refers to, it has a great deal of inexpensive environmental quality, while in *Rich*, on the other hand, environmental quality is costly and scarce. So unless the price of a unit of environmental quality is extremely high or extremely low, individuals in *both Poor* and *Rich* will gain from trade.

So if individuals were all rational and well informed, and it were possible for individuals easily to buy, sell, and transport pollution or "environmental quality," there would be active trading between the developed and less developed nations of the world, and pollution would be pouring out of the developed nations and into the less developed nations. But units of environmental quality cannot ("lamentably"?) be individually appropriated, bought, and sold, and it is hard to transport pollution between nations. Consequently, Summers argues, the World Bank can enhance world welfare by facilitating transfers of pollutants to LDC's in return for some measure of compensation.

Merely shifting pollution to LDC's, without paying any compensation would not, of course, be *mutually* beneficial. But from the perspective of cost-benefit analysis, it would still result in a net

benefit, because the developed countries *could* compensate the LDC's and still be better off.⁶ The benefits are larger than the costs. One could thus read Summers' memo as also pointing out the advantages of *uncompensated* transfers of pollution to the LDC's. It might seem unfair that the benefits go to developed nations and the costs to LDC's, but cost-benefit analysis is not a theory of fairness. It is, instead, an attempt to identify which policies provide net benefits. Most economists would say that issues of fairness need to be addressed separately.

Why should one then conclude, as Summers seems to, that it is "lamentable" that "pollution is generated by non-tradable industries?" Why should one conclude, "Clearly trade in goods that embody aesthetic pollution concerns could be welfare-enhancing?" How is one supposed to reach the conclusion that "the World Bank [should] be encouraging *more* migration of the dirty industries to the LDC's?" How does one get from claims about how rational and well-informed individuals *would choose* to claims about *welfare* and from claims about welfare to claims about what the World Bank *ought to do*? What is the logic of Summers' argument?

I suggest that Summers' argument can be spelled-out as follows:

- 1 For some intermediate amount of compensation C , all rational and well-informed individuals whether in developed countries or in LDCs would be willing to transfer pollution from a developed country to a LDC. (premise)
- 2 If all well-informed and rational individuals are willing to make an exchange, then carrying it out makes all of them better off. It increases everyone's welfare. (premise)
- 3 Shifting pollution to LDC's from developed countries and paying compensation C makes everyone better off (from 1 and 2).
- 4 One should do what makes people better off (premise)
- 5 One should shift pollution to LDC's and pay compensation C (from 3 and 4).

The conclusion in this reconstruction of Summers' argument is slightly different than "the World Bank [should] be encouraging *more* migration of the dirty industries to the LDC's," but if one assumes that the jobs and revenues provided by dirty industries are adequate compensation, then this reconstruction captures Summers' intentions.

The above argument contains two apparently plausible moral premises, statements 2 and 4. It is important to recognize that they are indispensable. Summers is well aware that dumping pollution in LDC's raises moral questions, but the tone of the memorandum (which is addressed to fellow

⁶ Economists would say that the uncompensated transfer of pollution to LDC's would constitute a "potential Pareto improvement." If compensation were paid, everyone would prefer the transfer, which would then be an actual Pareto improvement. See Hicks 1939 and Kaldor 1939. Changes that produce net benefits are potential Pareto improvements.

economists, not to the public at large) misleadingly suggests that the three numbered paragraphs make a “scientific” case, while the last paragraph in the memo mentions wishy-washy moral objections. The moral content does not however wait for the last paragraph to make its appearance. The three numbered paragraphs are part of a *moral* argument, and Summers’ memorandum is shot through with ethics.

The uproar caused by this memo suggests that most people are not willing to accept its conclusion. Why shouldn’t the World Bank encourage migration of dirty industries? One reason is that encouraging dirty industries to migrate to LDC’s might lead to more total pollution. Developed countries have stronger incentives and greater administrative capacity to enforce pollution controls than do LDC’s. Although this is a serious objection, it does not challenge Summers’ moral framework, and I shall say nothing more about it.

Second, even if people in both developed economies and LDC’s would be happy to shift pollution to LDC’s for some reasonable compensation, one might complain that the exchange is *unfair*. Developed countries are exploiting the poverty of LDCs – which, in addition, they are often responsible for. Summers alludes to such objections in his last paragraph, but he offers no response apart from the observation that such objections “could be turned around and used more or less effectively against every Bank proposal for liberalisation.” The injustice objection shows that premise 4, that one ought to do what makes people better off, needs qualification. It may not be right to make people better off if doing so involves injustice.

Is there no more fundamental objection to the whole mode of argument? Since there is no problem with the logic of the argument, if the conclusion is mistaken there must be some mistake in its premises. We have already seen that premise 4 needs to be modified. What about premise 2, that mutual willingness to exchange implies mutual benefit. It is comparatively easy to question this premise on the grounds that the parties may be ignorant or irrational. But as already noted, Summers argues that, given the *actual consequences* of increased pollution in LDC’s, rational individuals in LDC’s *should be* willing to accept pollution for less money than individuals in developed nations should be willing to pay. When mutual willingness to exchange is not based on any mistake, does it not then follow that exchanging would be mutually beneficial? In the discussion of preferences and well-being in the next three sections, I shall provide some reasons to deny this qualified version of premise 2, but for the moment let us suppose that it is acceptable.

That leaves premise 1, that all rational agents would agree to the exchange. Recall that this premise is itself the conclusion of an argument from the fact that the (economic) costs of pollution are lower in LDCs than in developed countries. But *rational and well-informed individuals do not have to accept the market’s evaluation of the consequences of the pollution*. Premise 1 is a moral

premise, too. The economic costs of the harms pollution causes are lower in LDC's because wages and productivity are lower, because people are more likely to die of other things before they can be harmed by some kinds of pollution, and because there are other pressing needs upon which individuals will spend their money first. But are economic costs and benefits a reliable guide to what is harmful and beneficial? Given the current unequal distribution of wealth, preventing or curing paraplegia confers much greater economic benefits in rich countries than in poor ones. But the moral significance of paraplegia should not depend on whether the victim lives in a wealthy country or on the victim's own wealth. One can thus reasonably raise moral objections to regarding economic costs and benefits as a guide to what *ought* to be done. Costs and prices have a contestable moral significance built into them.

Summers reduces the question of whether LDC's are "underpolluted" to the question of whether the welfare consequences of shifting more pollution to the LDC's would be favorable. "Welfare" for Summers, as for most economists, is preference satisfaction. The "cost" of the consequences of pollution is thus the amount by which people's preferences are less well satisfied. And Summers' measure of preference satisfaction is willingness-to-pay. A permanent injury to a child lowers the welfare of the child's family much more if the family happens to live in a rich country, because the cost of the child's care, the extent to which the child's prospective earnings are diminished, and the amount which the family would be willing to pay to avoid the injury are all much greater in rich countries than in poor. Measuring costs and benefits this way seems morally unacceptable. It is just as important morally to avoid crippling a poor child as to avoid crippling a rich one.⁷

This examination of Summers' memorandum makes obvious the central problem with valuing health states and interventions by willingness to pay – which is that goods and harms suffered by

7 Summers' other argument does not have this flaw. In some cases a given exposure to a pollutant will in fact diminish the health and welfare of people in LDC's less than it will diminish the health and welfare of people in rich countries. If, to use Summer's own example, a pollutant increases the risk of prostate cancer, which is a disease of elderly men, then the pollutant will not increase risk of suffering or death as much, if few men live long enough to contract the disease. Furthermore, for purely medical reasons, a given dose of a particular pollutant may have fewer negative health consequences if there is already very little pollution, than if there is a great deal. Although these differential effects might not exist if there were not other inequalities between developing nations and LDCs, claiming that such pollution has a lower cost in LDCs does not involve valuing the lives of those who live in LDCs less. The weight of these arguments can be questioned, however, since the consequences of increased pollution may lie many years in the future, when the differences in longevity and levels of background pollution on which the differences in effect depend may have disappeared. Furthermore, the interaction between pollution effects and the generally worse health status of people in LDCs might render some of the effects of pollutants more rather than less serious. The numbers of people affected by pollution must also be considered. Adding up all these factors, there seems to be no justified presumption that transfers of pollution toward poor countries is morally desirable. It is thus questionable whether thoughtful people should or would be willing to transfer pollution from developed to developing countries. One further problem should also be mentioned. The idea of compensating a *country* is a cheat: to claim that everyone would be willing to transfer pollution illegitimately treats countries as if they were individuals. Even in its own terms the argument does not go through, because the compensation may fail to reach the individuals who are harmed by the pollution.

those with little wealth and low expected earnings will count for less than goods or harms suffered by those who are wealthy and who have large expected earnings. (If the victims of sleeping sickness were rich, they would not have to depend on the fortunate coincidence that eflornithine removes facial hair in order to have access to this cure for the disease.) The economic costs of the death of a lawyer or engineer will typically be much larger than the economic costs of the death of an unskilled laborer of the same age. Yet people are not – and should not be – willing to accept the economic evaluation of their deaths as the proper evaluation to guide health assessments and policies. A country where morbidity and infant mortality are concentrated in the poor and unskilled will, other things being equal, have a higher level of welfare and a greater rate of economic growth than a country in which the same morbidity and infant mortality is evenly distributed among the poor and the wealthy. But it does not follow that the overall level of health is higher in the first country or that its health policy is better than the policy in the second country. Measures of health benefits in terms of willingness to pay consequently play a very small role in health economics (but see Olsen 1997).

3 Eliciting preferences for health states and interventions

Since, as we have just seen, it is unacceptable to measure preferences for health states and health improvements in terms of willingness to pay, either one must surrender the idea of valuing health states and interventions in terms of preferences, or one must find some other way of measuring preferences. Health economists, have for the most part taken the latter alternative and have attempted to measure preferences for health states directly or by examining trade-offs individuals would be willing to make between more life and healthier life.⁸

They have thus rejected cost-benefit analysis. In its place many defend “cost-effectiveness analysis,” in which different health improvements are compared in terms of the ratio of their expected benefits – measured in units such as years of healthy life – to the monetary costs. Switching from cost-benefit to cost-effectiveness analysis avoids the distortions introduced by measuring preferences in terms of willingness to pay, but it also prevents economists from comparing the benefits of devoting resources to health provision as opposed to some different activity. For example, cost-effectiveness analysis has nothing to say about whether resources are better spent on

⁸ This claim is controversial, and in a comment on an earlier version of this essay, Erik Nord challenged it. In his view, health economists have instead clung only to the terminology of “preferences” while in fact shifting to the position that the evaluation of health states should be informed by the *judgments* of the members of the target population. For a discussion of the distinction between preferences and judgments, see the beginning of section 4.

health or education.

In this section I shall comment briefly on methods of measuring preferences that do not rely on willingness to pay, because, as we shall see, the problems with the specific methods – which are numerous and only some of which I shall mention – link up to fundamental difficulties in the interpretation of the value of health states as the extent to which they satisfy preferences.

The easiest way to investigate people's preferences among health states is just to ask them. But merely finding out how many people say they would prefer to be blind rather than to be deaf and how many would prefer to be deaf rather than blind provides no measure of the intensity of these preferences and no way to add up the various health states in order to determine some overall measurement of population health. So instead of asking, "Which do you prefer?" individuals are asked to rate health states on a zero to one-hundred scale or to locate them along "a visual analog scale" running from best to worst health states. Although this is a method of assigning numbers to health states, there is good reason to doubt whether the numerical assignments tell one more than simply the preference ordering (Nord 1999, pp. 91-94).

Three other methods of valuing health states have been employed, which are more complicated than simply asking individuals to rate the states: the standard gamble, the time trade-off, and the person-trade off.⁹ In the standard gamble, individuals are asked to name a probability p such that they would be indifferent between continuing in a particular health state (such as being blind) for a year and facing a gamble that will give them perfect health for a year with probability p and death with probability $(1 - p)$. If individual's preferences satisfy the axioms of expected utility theory¹⁰ and one assigns a utility of "1" to full health and "0" to death, then p is a cardinal measure of the utility of the diminished health state, which can be added up to determine overall health states.

In the time trade-off method, instead of being faced with a gamble, individuals are asked in effect what percentage of a year of perfect health would be equivalent to a year of life in a particular diminished health state, such as being deaf. That percentage, just like the probability p in the standard gamble, is a measure of the value of the diminished health state. One problem with both the standard gamble and the time trade-off method is that minor health problems may count as no health

9 For some technical criticisms of these methods, see Loomes and McKenzie 1989. In unpublished work, Costa and Rovira show that one can elicit preferences for health improvements by asking individuals hypothetical questions about which health programs they would fund out of a fixed budget. The authors do not address the question of why those who face the actual task of deciding how to allocate health care resources should be concerned to emulate these hypothetical decisions. Is there any better basis for taking the hypothetical decisions as determining the values of health improvements than for taking the actual decisions as doing so? If the question is how to make such decisions, the answer is not to cite what decisions have been made.

10 These require continuity, independence, and some technical conditions such as reduction of compound lotteries, in addition to completeness and transitivity.

problems at all. Many people claim that they would not trade a year of life with back pain or even with cancer symptoms for anything less than a full year of life in perfect health (Nord 1999, pp. 106-7). Similarly many people would not trade a year of life with a limp for a gamble that could give them a year of life without any disability but which has a one in a thousand or a one in ten thousand chance of immediate death. But obviously a population where people are suffering back pain or cancer symptoms is less healthy than a population without these problems. This last claim implies that either health is not determined by preference, or neither the time trade-off nor the standard gamble accurately measures preferences.

Still another method asks individuals to compare interventions that save the lives of some number of disabled individuals to interventions that save the lives of some smaller number of people without any disabilities (Patrick *et al.* 1973). Alternatively, one can ask individuals to compare an intervention that, for example, cures blindness in 1000 people and gives them a year of sighted, healthy life with another intervention that permits some smaller number of healthy people to live an extra year. The value of a year of life without sight can then be calculated.¹¹ Although the WHO has used variants of person trade-off methods such as the last one to calculate summary measures of population health, person trade-off preferences or judgments are sensitive to distributional concerns and to specific moral principles. For this reason, Nord has argued that person trade-offs should not be used to estimate the well-being of individuals in those health states, which Nord would measure in other ways. Some kinds of person trade offs are, he maintains, nevertheless extremely informative and should be used to elicit the weights society wishes to place on different health states for the purposes of allocation of health services.¹² Other kinds of person trade-off questions are inappropriate for both the purpose of measuring well-being and for the purpose of determining societal weights. For example, when asked the person trade-off question, "How many lives of paraplegics would have to be saved by a health care intervention in order for you to rank that intervention as highly as one that saves the lives of 1000 people without disabilities?" respondents might reasonably answer "1000" on the grounds that everyone's life is equally important. Such an answer does not imply that respondents believe that paraplegia makes no difference to the quality of life or that a treatment that leaves an individual a paraplegic is of the same value as one that restores an individual to full health.

All of these measurement methods have serious difficulties. People can only rank health states if

¹¹ Let x be the number of people and y be the value of a year of life when one is blind. Then the difference between a year of healthy life (=1) and a year of life while blind, $1 - y = x/1000$, or $y = (1000-x)/1000$.

¹² For a critique of Nord's proposal and a suggestion of an alternative approach, see Dolan 1998.

they understand what those health states are. But it is difficult to provide people with that knowledge and questionable to rely – as has often been the case – on the rankings provided by health-care workers. People often have trouble understanding the comparisons they are asked to make, and their answers may reflect these difficulties. The answers are sensitive to how the questions are framed: an intervention that leaves some people in a group to die may appear very different than one that saves the lives of the others. One finds many internal inconsistencies in the comparisons people make, and the answers people give one week may be very different than the answers they give the next.

Because of the problems concerning framing, instability, and inconsistency, surveys and polling are less often used, except for purposes of exploration or suggestion. Both the WHO and local organizations attempting to value health states increasingly make use of deliberative groups. Consider, for example, the WHO's 1996 protocol for measuring severity weights. Though the specific person-trade off methods have been abandoned, the protocol is still a good example of how deliberative groups are supposed to assist in the measurement process.¹³ Groups of 8-12 health-care workers were assembled and confronted with a description of a health-state whose severity was to be measured. Each member of the group was asked individually two person-trade-off questions. The second (PTO2) was similar to the person-trade-off question described above. The first (PTO1) was described to individuals as follows:

You are a decision maker that has only enough money to buy one of two mutually exclusive health interventions. If you purchase intervention A, you will extend the life of 1000 healthy individuals for exactly one year, at which point they will all die. If you do not purchase intervention A, they will all die today. The alternative use of your scarce resources is intervention B, with which you can extend the life of n individuals with a particular disabling condition for one year. If you do not buy intervention B, they will all die today; if you do purchase intervention B, they will all die at the end of exactly one year. For example, you may be faced with a choice between extending the life of (A) 1000 healthy individuals for one year, or (B) 2000 blind individuals for one year.... (Murray 1996, p. 91)

If one assumes that these two person-trade-off questions are merely alternative ways of measuring the severity of the disability, then the answers should be consistent. For example, if individuals say that they are indifferent between saving the lives of 1,500 blind people and saving the lives of 1,000 people without any disabilities, then they should be indifferent between enabling 3,000 of the blind to

¹³ I will say a little about the new protocols, which are still under development, near the end of the next section.

have a sighted year and extending the lives of 1000 disability-free individuals.¹⁴ But in fact individual's answers to these two questions are inconsistent. Individuals are then asked to reflect on their values until they come up with consistent answers to the PTO1 and PTO2 questions. After each individual has done so, the answers are shared with the other members of the group. "Those with the highest and lowest evaluations are asked to explain their *reasoning* to the group, and an open discussion follows" (Murray 1996, p. 92; my italics). Individuals can then privately adjust their evaluations.

This process was employed for 22 "indicator" health conditions. The evaluations of all other health states was inferred from the implicit weighting of different aspects of health states revealed by the evaluation of the indicator states. After all 22 indicator conditions were evaluated, group members were asked independently simply to rank the conditions and were asked to revise their ordering or their evaluations if the two were inconsistent. Murray reports that the WHO staff who worked with these groups felt that it was important to have roughly equal numbers of men and women, and they also believed that the group leaders or "facilitators" had an important role in encouraging the groups and getting them to remain serious (1996, p. 93).

Some of the features of this protocol can be seen as attempts to address specific objections to identifying well-being with the satisfaction of preferences. First, by having the group members be health-care professionals, one lessens the problems caused by ignorance and false beliefs. Problems caused by false beliefs are, however, bound to remain; because not all of the indicator conditions will necessarily be well understood by every health-care professional, and health-care professionals can in some cases be unaware of the difficulties health states cause patients in their day-to-day life.

Second, asking subjects to make the two different PTO judgments and to make their evaluations and rankings consistent seem intended to overcome some of the difficulties due to framing. The WHO protocol assumes that both PTO questions are probing for the same preferences for different health states and that by demanding consistency between the two judgments, one achieves a more accurate statement of the preference than by asking either of the person trade off questions separately. Either preferences are buried deeply within us and are hard to access, so that a variety of ways of eliciting them helps us to uncover what they are, or else – as seems more likely – people's "preferences" are not well articulated and the different methods of elicitation lead people to construct them in a more robust and defensible way.

¹⁴ If someone is indifferent between saving the life of 1,500 blind people and 1000 disability-free people, then the value of a year life if one is blind is 2/3. So the difference between a year of healthy life and a year of life with blindness is 1/3, which is the ratio between extending the life of 1000 healthy people and giving 3000 blind a sighted year.

Third, preferences between health states apparently depend upon reasons and arguments. In Murray's own words, "Those with the highest and lowest evaluations are asked to explain their *reasoning* to the group, and an open discussion follows" (Murray 1996, p. 92; my italics). Unfortunately there are no transcripts recording such reasoning or the ensuing discussion, and indeed I have been unable to get hold of any transcripts of any groups engaged in evaluating health states. So I can only guess about the nature of the reasoning. Much of it probably involves pointing out the consequences of different health states for common aspirations or occupations. Some of it may consist of reminders concerning features of the health states themselves and what it must be like to experience them. In any event it is obvious that the "preferences" being elicited here are unlike mere matters of taste. They are sensitive to reasoning and discussion, and it is assumed that the preferences individuals expressed after discussion are in some sense more "accurate" than those expressed before discussion.

4 Eliciting judgments versus measuring preferences or subjective states

Why might health economists prefer to rely on deliberative groups to evaluate health states rather than to carry out polls or surveys? One answer might be that polling is an unreliable way of measuring preferences. The results of polling are just too unstable, flimsy and inconsistent. Health economists might turn to deliberative groups for a better way of measuring preferences. In relying on them, economists might hope to overcome problems of ignorance and mistaken belief, to mitigate framing effects, and to arrive at more stable preferences.

But there is another and completely different diagnosis of the reliance on deliberative groups: Perhaps they are superior to polling or surveys because economists are not after all trying to measure preferences. Deliberative groups might be superior because they are better ways to elicit *judgments* or *opinions* concerning the severity of health states.

To clarify this distinction consider the following three questions concerned with evaluating health states:

1. Which health state *feels* better or worse?
2. Which health state is *preferred*?
3. Which health state *is* worse?

Since question one concerns how people feel, investigators might attempt to answer it by asking people, "Which would feel better or worse to you?" To investigate question two, one needs to study people's preferences. In principle, one could just ask, "Which health state would you prefer?" Question three, is not on the surface at least a question about feelings or preferences. There seems

nothing contradictory in saying that one health state is worse than another, even though the latter *feels* worse. Nor is there apparently anything contradictory about preferring the worse health state or even preferring a state of affairs in which one is, all things considered, worse off. Questions 3 thus appears to be distinct from either question 1 or question 2. If one were to try to answer question 3 by asking people, all one could ask for is their opinion. The contrast between preference and opinion is the contrast between answers to question two and answers to question three, while question one concerns mental states or feelings rather than either preference or judgment (though its answer may provide reasons for preferences and judgments).

If the evaluation of health states depended entirely on how those health states feel, then those concerned with population health should be asking questions of the first kind. In just the same way, if one wants to know which of two foods taste best to the members of some population, the results of taste tests are decisive. One need only measure subjective reactions. If the evaluation of health states were like this, then question one would be the right question to ask, and the task would be to measure subjective reactions.

It is not easy to measure subjective reactions. The only way to do so is to study how people feel. If asked whether the felt quality of life is worse with depression or with paraplegia, most people would not be in a good position to answer. Researchers might seek out people who have separately experienced both these health states, but there are not many of them, and memories of how health states feel are not very reliable. Alternatively, one can poll people who are in these health states concerning how they feel, though those who are depressed may be unreliable sources. Since different people may give different responses, investigators would also need some method for determining an average or typical response. The central point is that the way to answer questions of the first kind – that is, questions concerning the felt quality of life in different health states – is to measure people's feelings. The difficulties in answering question one are measurement problems.

One cannot evaluate health states accurately by measuring subjective states, because the felt quality of life is obviously not the only factor upon which the evaluation of health states depend. The evaluation of health states also depends on more objective factors, such as functional limitations. Furthermore, as I have argued above, economists claim to be concerned with the extent to which people's preferences are in fact satisfied rather than with their subjective states. Although measuring preferences is obviously not easy, it may be easier for ordinary people to express preferences among health states such as depression and paraplegia than for ordinary people to know how to compare the felt quality of life in those states. Although one's preference might change dramatically with better knowledge of the health states, ignorance is no barrier to having a preference. There is thus no reason why people might not say that they would prefer to be depressed

rather than crippled, even though they do not know how it feels to be either. And someone might, of course, prefer depression to paraplegia, even knowing that it feels worse, because mobility is more important to this person than is the way life feels.

If the evaluation of health states were a matter of preferences – if those who were attempting to evaluate health states were asking question two – then the evaluation of health states would depend entirely on the measurement of preferences. If one health state were better than another if and only if on average people prefer the first to the second, then the difficulties in evaluating health states would only be measurement difficulties; and there would be no role for ethical reflection. People’s preferences – their answers to question two – might, of course, be distorted or faulty. They might reflect inconsistent weighting of different dimensions of health states. Health economists might be convinced that the preferences they measure would change dramatically if people were to reflect more carefully or to have more experience or knowledge. But regardless of these distortions or instabilities, if one health state is better than another if and only if people to prefer it, then measuring preferences would determine the evaluation of health states. Question two would be the only question to ask.¹⁵

But, as I shall argue below, it is not true that one health state is better than another if and only if people prefer it. If a health state lowers expectations and aspirations, then preferences may be better satisfied, but lower expectations and aspirations do not imply better health. Preferences (and feelings) matter, but they do not *constitute* the evaluation of health states. They are not decisive. If one wants to know which of two foods the population would prefer that the government distribute to the hungry, then the results of polling are decisive. One need only measure preferences. (These preferences would depend on tastes, but also on beliefs concerning, for example, which food was more nourishing.) If the evaluation of health states were like this, then question two would be the right question to ask, and the task would be to measure preferences. There would be no question concerning whether the preferences are correct or incorrect representations of something else. The question would *concern* the preferences, and the preferences would constitute the answer.

The evaluation of health states is not, however, like the comparison of the taste of foods nor the comparison of popular support for distributing one food or the other. It is instead, a matter of judgment, like the question of which food would be better to distribute. Preferences for one policy or one health state over another may be unanimous, rational, and well-informed; but nevertheless mistaken. Whether one health state is better than another *depends* on feelings and preferences, but

¹⁵ Those who maintain that the evaluation of health states depends on informed or rational or “laundered” preferences, rather than on actual preferences might want to ask question 1 as well, because information about the felt quality of life might be needed in order to have an informed preference.

it is not constituted by them and cannot always be settled by them.

The difference I am pointing to in distinguishing *preferences* between x and y from *opinions* concerning which is better is principally a contrast between *constituting* and *assessing*. If health state x is better than health state y if and only if people prefer x to y , then measuring the preferences determines the ranking. The preferences *make it the case* that x is better or worse than y rather than reporting on what is independently the case. This is true regardless of whether one holds the view, which is common among economists, that well-being is the satisfaction of actual preferences or whether one holds the view which is more common among philosophers, that well-being is the satisfaction of “informed” or “rational” or “laundered” preferences. The only difference concerns *which* preferences determine what is better or worse for me. According to those who take preferences to determine the evaluation of health states, investigating the severity of health states by measuring people’s preferences is basically just like investigating the prevalence of hypertension by measuring people’s blood pressure.

To take preferences to constitute and determine the ranking of health states does not presuppose that preferences are infallible or incorrigible. Preferences can be misreported and mismeasured, and individuals can sometimes be mistaken about what their preferences are. Furthermore, when people learn more or acquire more experience their preferences may change. What is infallible is instead the *relationship* between preferences and evaluations. If one has measured the right preferences and has measured them correctly, then the health state evaluation is determined. If one knows the relevant preferences, then one cannot be wrong about the health state evaluations. One has a kind of conditional infallibility. The infallibility of the inference from preference to evaluation contrasts with the fallibility of the inference from opinion to fact.

In the judgment that x is better than y , the ranking of x and y is assessed rather than determined. Given knowledge of the relevant judgments or opinions, one can still be wrong about how to evaluate the health states. Evaluating health states by eliciting judgments is more like investigating a crime by interviewing witnesses than it is like measuring blood pressure to determine the prevalence of hypertension. No matter how carefully considered, a population’s consensus that x is a better health state than y does not make it so. If those who are trying to evaluate health states are, as I shall argue in this section, studying judgments as well as preferences, then, as I shall argue in the next section, what they learn will not necessarily settle the evaluative questions. To know that most people think that health state x is worse than health state y may be of interest, especially if one also finds out people’s reasons. But, as I shall argue in the section 6, it no more settles the evaluation, than the opinion of the population concerning policy alternatives determines which is best.

Although the words, “preference” and “judgment” suggest the contrast I am drawing, in everyday

life people may speak of “judging” which cake one likes best (which is a matter of preference), or of “preferring” one answer to a factual question to another (which is a matter of judgment). Consider the following remarks of Robert Kaplan and John Anderson, “The empirical means of accomplishing this is measured *preferences* for the health states. These might be regarded as ‘quality judgments’ (1988, p. 211 [their emphasis]). Taken out of context, there is no way to determine whether Kaplan and Anderson are concerned with preferences or judgments in the sense that is relevant here.

In the remainder of this section, I shall argue that health state evaluations are not constituted by preferences. They are judgments. Information concerning preferences and feelings is relevant to these judgments, but preferences and feelings do not settle the evaluation by themselves. My argument will proceed in two stages. First I shall show that – regardless of the terminology they have employed – those concerned with evaluating health states either for the purposes of measuring population health or for the purposes of health care allocation have been concerned to elicit judgments and to measure feelings, as well as to measure preferences. Then I will argue that the evaluation of health states is a matter of judgment and that health economists have consequently been right to be concerned to elicit judgments. The fact that the evaluation of health states is a matter of judgment does not imply that measuring preferences and subjective feelings is irrelevant. On the contrary such measurement has a legitimate place in the process of evaluating health states, because information about feelings and preferences for health states is relevant to judgments about their value. In this section, my quarrel with some health economists is only that they have conflated the measurement of preferences with the measurement of feelings and the elicitation of judgments. After arguing in the next section that health states evaluation requires more than measuring preferences, sections 6 and 7 will turn to questions concerning whether eliciting judgments can answer ethical questions concerning health and health systems and eliminate any need for ethical reflection.

There are five features of the protocol the WHO has used that show that the WHO (like others) has been concerned to elicit judgments and to measure feelings rather than only to measure preferences. What has been called “the measurement of preference” is often not that at all. It is instead sometimes the measurement of feeling and more frequently the elicitation of judgment.

(1) Notice that the WHO protocol used groups consisting of health-care professionals, rather than a representative sample of the target population, on the grounds that health-care professionals were more *knowledgeable* concerning the indicator health states. The presumption is that it is better to rely on evaluations made by those who are more knowledgeable. Why? If one is measuring preferences, why measure the preferences of those who know more? What is “better” about their preferences? One might reply that the preferences of those who know more are more stable or have

less individual variation, or that those who are more knowledgeable are better able to report their true preferences. But if the task were to measure population preferences, it would be strange to rely on such an unrepresentative sample. A more plausible answer is, I think, that the judgments of health-care professionals concerning the severity of health states are likely to be more *accurate* than the judgments of those with less knowledge and experience. Once one admits the thought that there are more or less accurate comparisons of health states, one has admitted that the evaluation of health states involves the judgment of something “external” – of a value that is not constituted or determined by the assessor’s preferences.

(2) The WHO is already committed to *overruling* some preferences. There are, for example, cultures in the world today that would treat diseases affecting women as less serious than diseases affecting men, because they value men’s health more highly than women’s health. Even if such evaluations survived criticism and discussion (though one hopes, of course, that they would not), the WHO would ignore such evaluations in determining the severity of health states and in measuring overall population health. If the evaluation of health states were simply a matter of measuring preferences, such overruling of preferences would be illegitimate. Some speak of measuring “laundered preferences” rather than actual preferences (Murray 1996, pp. 4-5; citing Goodin 1986, pp. 75-77), but if laundering requires overruling preferences, even when preferences survive criticism and discussion, then it seems that the evaluation of health states cannot be determined by preferences.

(3) Questions related to the dependence of preferences on knowledge lead to questions concerning *which* preferences or *whose* preferences should count. I already touched on this point about in (1), when I pointed out that if one wants to measure preferences, one should seek a representative sample, while if one seeks to elicit reliable opinions, then one would look for respondents whose opinions are more likely to be correct. But some of the discussion concerning whose preferences to measure suggest that economists have been concerned to measure feelings rather than preferences. As is well known, the evaluations of health states, especially of continuing disabilities, differ dramatically depending on who is doing the evaluating, with those who have been disabled for long periods adapting to their disabilities and ranking their health states as little worse than those who are not disabled. In these cases, there appear to be changes both in individual’s preferences and in individual’s judgments concerning the severity of health states.

Why do the preferences and judgments of those who experience disabilities change? Is it because they come to have better knowledge of what life with the disabilities is like, or do their preferences change for some other reason? This question may be of importance even to those who believe that the evaluation of health states is merely a matter of measuring preferences, if they believe that only

certain kinds of preferences count. But one can also explain why the explanation of the change matters if one takes the health economist as concerned with the judgment of how bad certain disabilities are and with how life feels with disabilities, rather than merely with the strength of people's preferences for avoiding those disabilities. If preferences differ, because those who are disabled have better knowledge of what it is like to be disabled, then the evaluations of those who are disabled should be decisive, on the grounds that the evaluations of those who know more are likely to be more accurate. The evaluation depends on knowledge, and it is not settled by preferences.

Consider another possible explanation for the differences in evaluations. Murray writes, "The fact that some slave may have been happy does not in any way make the reality of slavery more acceptable. By analogy, the adaptive powers of man should not make the prevention or rehabilitation of those with a non-fatal health outcome less valuable" (1996, p. 31). If the evaluation of health states were determined by feelings or preferences, then what Murray claims would be mistaken. If the extent to which a disability diminishes health is constituted by how one feels or what one prefers, then those who are blind and contented or who are blind and have no strong preference to see again are not much less one healthy than those who are sighted, and curing their blindness would not increase population health significantly. So Murray must be denying that health state evaluations are determined by feelings or actual preferences. He is suggesting instead that feelings may be a bad guide to health and that actual preferences can sometimes be distorted and mistaken (see Elster's 1983 discussion of "adaptive preferences").

Murray is thus denying that the evaluation of health states is determined by actual preferences. He must instead maintain either that they are determined by "informed," or "laundered" preferences or that they are not determined by preferences at all. He takes the difficulties posed by adaptive preferences as an argument for focusing on laundered rather than actual preference, but his qualms are also consistent with the denial that health state evaluations are constituted by any preferences, no matter how cleansed.

Nord grants Murray's view that health-care priorities should not be determined by actual preferences, but he argues that distortions in preferences are irrelevant to the *utility* of a health state, "Coping affects utility and should therefore be incorporated when utility measurement is the issue" (1999, p. 89). If Nord means to maintain that distortions should never lead us to disregard preferences, when we are measuring well-being, then his point is exaggerated. But he goes on to make a further observation, which counts strongly against identifying health care evaluations with feelings or preferences, whether informed or not. "The argument reveals a more basic problem.... Society may very well wish to take into account *both* objective symptom relief and functional improvement *and* the increase in subjective utility (quality of life) when valuing a health service

(1999, p. 89). Nord is claiming that evaluations of health states depend on *both* “subjective utility” and “objective symptom relief and functional improvement.” So the evaluation of health states is not fixed by feelings or preferences.

In the passages quoted above, both Murray and Nord have shifted to a conception of utility or well-being as a mental state rather than as an index of preference. So Nord is saying is that the comparison of health states does not depend exclusively upon the comparison of the mental states of those who occupy those health states. Those who are concerned with health state evaluation must ask other questions besides question one. Nord could, however, make the same point concerning preference satisfaction: a comparison of the extent to which their preferences are satisfied (the answer to question two) is not decisive either, because the severity of health states also depends on relatively objective factors. Even if welfare were the satisfaction of preferences (which it is not), the evaluation of health states does not depend only on the well-being associated with those health states.

(4) The multiplicity of different questions that are involved in the evaluation of health states provides further evidence that health economists are frequently eliciting judgments rather than measuring preferences. In the WHO protocol discussed above, individuals were asked to answer both PTO1 and PTO2 questions and to make consistent the evaluations elicited in these two ways. Nord argues that this is a mistake. In his view, the first question – saving x blind versus 1000 healthy – ought to trigger the moral commitment to the equal value of all lives and that consequently the answers to the two PTO questions *should be* inconsistent (see also Harris 1987). When the subjects are told, however, “For example, you may be faced with a choice between extending the life of (A) 1000 healthy individuals for one year, or (B) 2000 blind individuals for one year.... (Murray 1996, p. 92), they are discouraged from thinking that saving the life of someone with a disability is just as important as saving the life of someone who is healthy. Nord believes that in forcing consistency the protocol drives people to make compromises between competing considerations and that the numbers that result have little to do with the severity of disabilities. Whether Nord is right or not, it is obvious that both the interpretation of the question by the respondents and the interpretation of the answers by the health economists requires reflection on what moral principles apply. And insofar as the answers depend on moral principles and on decisions concerning whether moral principles apply, they reflect judgments.

The question of whether an extra year of life for someone who is disabled has the same “value” as an extra year of life for someone who is not disabled in fact involves important ambiguities, which lead Murray and Nord to talk at cross purposes (Anand and Hanson 1997). Nord is concerned to measure the social value of health care interventions so as to distribute health care resources to the

most valued uses. Murray is concerned with measuring health. They are not measuring the same thing, unless the only thing that matters in a health-care intervention is the quantity of health benefits. But, as Nord documents (1999, esp. ch. 4 and Nord *et al.* 1999) and Murray concedes (Murray and Lopez 1999, p. 18), people have moral commitments that conflict with maximizing health benefits. They show a special concern with those who are seriously ill. They are concerned to distribute health-care resources so that even those who will profit less from treatment can have a chance of being treated. They believe that saving a life ought to count equally, whether someone is disabled or not.

These are separate considerations than quantity of health. Suppose for example that a population faces an epidemic that will kill everyone. Half the population is disabled. There are two ways to combat the epidemic: *A* gives everybody a 50% chance of survival, while *B* saves the lives of all those who are not disabled. Given the social values that Nord documents, intervention *A* is morally superior to intervention *B*, because it gives everybody an equal chance of surviving and counts saving the life of someone who is disabled as just as valuable as saving the life of someone who is not disabled. But the two interventions could not possibly make an equal contribution to the *health* of the population, since the population that survives after intervention *B* has no disabilities, while half the population that survives after intervention *A* is disabled. From the perspective of the attempt to define a summary measure of population *health*, intervention *B* must bring more health than intervention *A* – regardless of whether it is otherwise morally acceptable.

Instead of preference for one health state compared to others, which can be measured in a variety of different ways, one has here different evaluative judgments. It is as if individuals are being asked “Which would you ‘prefer’ in terms of health benefits?” and “Which would you ‘prefer’ in terms of justice?” But “prefer” here is a synonym for “judge as superior.” Preferences do not *constitute* the answer to either question. One can use the words, “Which do you prefer, *A*, or *B*?” but one is asking for people’s answers to specific questions concerning which provides the greater health benefits or which is fairer. Their answers express *opinions* rather than preferences for *A* or *B*.

(5) The WHO protocol’s insistence on the need for reasoning and discussion provides further evidence that they are eliciting judgments as well as or instead of measuring preferences. Without evidence concerning the precise nature of the discussions within deliberative groups, this point is speculative, but the importance of reasoning points toward evaluations being a matter of judgment and dependent on reasons. Nord points out that when person-trade off questions explicitly indicate ethical positions, subjects adopt different responses than when they are left to conceive of prioritization problems themselves. Consider what Nord concludes,

...it would be naive to think that it [the person trade-off approach] is an easy way to obtain

accurate numerical representations of societal preferences for resource allocation....Some sensible measurement strategies are already indicated at this point: The technique needs to be applied in fairly large groups of subjects to keep random measurement error at an acceptable level and to avoid political biases. To control for framing effects, it seems important to take subjects through a multistep procedure, in which they are induced to consider carefully the various arguments that might be relevant in each comparison and to reconsider initial responses in the light of their implications. In other words, the investigator ideally should be seeking to establish a “reflective equilibrium” (Rawls 1971) in his or her subjects....It is of course very difficult to fulfill all these requirements using self-administered questionnaires; direct, personal communication seems necessary. The individual interviews reported by Rosser and Kind (1978) are an example of how this may be done. Seminars described in Nord (1994) and Murray and Lopez (1996) are other examples. (Nord 1999, p. 131)

Nord regards the person trade-off approach as a method to obtain “numerical representations of societal *preferences* for resource allocation [italics added].” Unfortunately there are many sources of measurement error, and steps must be taken to minimize those errors. To cope with the difficulties posed by framing effects, Nord suggests that “the investigator ideally should be seeking to establish a “reflective equilibrium” (Rawls 1971) in his or her subjects” and doing so requires individual interviews or “seminars” (which I have been calling deliberative groups). Recall that Rawls’ notion of a reflective equilibrium involves a dialectical process of modifying specific considered moral judgments, moral principles, and hypotheses concerning the nature of morality until they are consistent with one another and mutually supporting (1971, pp. 20, 48-51; see also Daniels 1979). Although one can describe a process of modifying preferences to make them consistent with one another, it would be odd to describe the results as a reflective equilibrium, since there seems, for example, no good analogy to the relationship between moral principles and specific moral judgments. While ostensibly discussing problems in the measurement of preferences, Nord is instead concerned with how to help people to make good *judgments* about how to allocate health-care resources.

5 Why Measuring Preferences Is not Enough

In the last section, we considered a number of features of health state evaluations that indicate that economists have been concerned to elicit judgments or opinions and to measure subjective states rather than only to measure preferences. The 1996 WHO protocol treats health state evaluations, like judgments and unlike preferences, as correct or incorrect, and investigates the “preferences” of

experts rather than attempting to sample the preferences of the target population. In revising these protocols, however, the WHO seems to be moving toward a preference-measurement model of health-state evaluations and away from a model that bases health-state evaluations on opinions of the target population. In this section I shall argue that this change of course is in many ways a great improvement, because preferences do matter; and eliciting opinions should not be confused with measuring preferences. But I shall also argue that measuring preferences is not enough to determine how to evaluate health states.

For health state valuations in its current project (Global Burden of Disease 2000), the WHO is undertaking a two-tiered data collection strategy that consists of both surveying the general population and working with groups of educated respondents.¹⁶ Representative samples of the general population will be polled concerning their ordinal rankings of health states, and they will be asked to locate health states along a 0 – 100 visual analog scale. In this way the WHO will address questions concerning how preferences vary among groups and across nations. Groups of educated respondents will be asked to rank health states using versions of the standard gamble, time trade-off and person trade-off methods, in addition to being asked to rank health states and to locate them along a 0-100 visual analog scale. The version of the person trade-off method to be used is basically what was called above PTO2, and so individuals will never be asked the person trade-off question that collides with the principle that everyone's life is equally worth saving. The reason for relying on groups of educated respondents is that uneducated respondents have great difficulty with more complicated elicitation techniques, while location of health states along a visual analog scale does not provide a reliable interval measurement of preferences. The responses of educated informants enables the WHO to translate responses obtained using the visual analog scale to an interval-scaled measure of "strength of preference."

The reason why preferences are measured in several different ways is that each individual method is biased or distorted.¹⁷ The visual analog scale winds up providing only ordinal information. The standard gamble is biased by risk aversion. The time trade-off method is biased by time-preferences. The person trade-off method is biased by ethical concerns. Just as one can edge one's way up a chimney by pressing on some of the walls while advancing along another, so one can calibrate the responses to each of these different questions against one another and infer (fallibly, of course!) the

16 No account of these methods has yet been published. The description I give here derives from information that Joshua Solomon was kind enough to provide in a personal communication.

17 Why not make use of willingness to pay information too? Of course willingness to pay is a distorted measure of preferences, but so are standard gambles (by attitudes toward risk), time trade-offs (by attitudes toward the future), and person trade-offs (by moral commitments). If anything, it would appear that the distortion in willingness to

true strength of preferences for different health states. As before, the WHO plans directly to measure preferences for only a small number of health states. These preferences will enable the WHO to infer preferences along seven different dimensions or domains of health (which consist of mobility, self-care, usual activities, pain and discomfort, cognition, anxiety and depression, and social participation). One can then impute preferences for all other health states by locating them along these dimensions.

These methods appear to have a better claim to be measuring preferences. The use of educated respondents is justified without any suggestion that educated respondents are more knowledgeable. By surveying representative samples, the WHO can claim to be measuring population preferences. Although the educated respondents will work in small groups, and there will be discussion of the tasks, the plan calls for individual completion of each task, without the discussion and argumentation built into the older protocols. Rather than calling on the respondents to make their responses to the separate elicitation methods consistent, the responses are treated as raw materials to be corrected by the WHO analysts.

In one important way this return to measuring preference is a good thing. As I noted before, measuring preferences should have an important role in the process of evaluating health states, because information about feelings and preferences for health states is relevant to judgments about their value. *One* of the things that makes one health state better than another is that people prefer it. So measuring preferences plays an important role, and insofar as the new protocols enable the WHO to measure preferences rather than to elicit opinions and misleadingly to call them preferences, the changes are a very good thing.

But it is nevertheless still the case that health state evaluations cannot be based *entirely* on the measurement of preferences. The WHO is still committed simply to overruling some preferences. Health states can be evaluated or judged in several different ways, while it seems that there is a single preference relation. As Nord points out, some kinds of health state evaluations – especially those that are relevant to the allocation of health-care resources – should not be based exclusively on the well-being of people in those health states. Finally, health state evaluations, as is typical of judgments depend on reasoning and deliberation.

These considerations provide, I believe, conclusive reason to deny that the comparative evaluation of health states is *constituted by* preferences for those health states. The severity of health states does not depend only on preferences, but also on objective features. No matter how well laundered, “preferences” can be mistaken – they do not constitute the truth the WHO is looking for. Since the comparative evaluation of health states is not constituted by preferences, it should come

pay introduced by wealth would be easier to measure and correct for.

as no surprise that the evaluation is not determined by the measurement of preferences.

Health does not depend only on well-being, and well-being is not the satisfaction of preferences. The fact that preference satisfaction *is relevant to* health and well-being does not imply that preference satisfaction *constitutes* either of these. Even if one takes the comparison of very different health states to be a comparison of the average well-being of people who are in those states of health (as opposed to some more objective comparison), there is still an enormous role for rational argument and investigation. Of which goods are those with paraplegia typically deprived? Of which goods are those who are depressed typically deprived? How important are those goods? How easily are other goods substituted for them? At some point reasons may give out (though this is not obvious), and the assessment may end with the unexplained fact that members of the target population (or perhaps all people) simply do (on average) value some goods more highly than others. But such “facts” concern the evaluation of whole aspects of life, not preferences for health states, and the evaluation of health states consequently cannot be reduced to the measurement of preferences.

In a comment on an earlier version of this essay, Robert Goodin made a suggestion that helps to reinforce the argument that the evaluation of health states cannot be determined by measuring preferences. He maintained that the methods employed ostensibly to measure preferences are not measuring preferences at all. Instead they are eliciting *predictions* of what preferences *would be* if one were in the health states being compared. Evaluating health states in this way is not analogous to inquiring about whether eating salty foods raises blood pressure by taking people’s blood pressure on different diets. It is instead analogous to asking people whether they think their pressure would be lower if they ate less salt.

Goodin’s observation is however less conclusive than it may appear. If one insists that *preferences* rather than *feelings* are what matters, then there is no reason to deny that people can express preferences concerning health states they have not experienced. Fortunately at this moment I cannot feel what life is like when one is depressed or when one is deaf. If ask to compare how they feel, I can only predict how life would feel. But I can express a preference between these health states right now, without being able to compare how they in fact feel. Of course that preference ranking might change if I had better knowledge of what my life would be like if I were disabled or ill, but my current preference need not be a prediction of what my preference would be.

Goodin’s argument is, however, still potent, because the fact that I can express a preference ranking now of health states I have not experienced invites one to question whether that preference should matter. If my current preference for paraplegia versus depression differs from what my preference would be if I experienced separately both these states, then it would appear that my

current preference is not a good guide to the severity of these health states. My preferences for paraplegia versus depression are likely to be different, depending on whether I am depressed, or paraplegic or have been cured of both these conditions. *None of these preferences is definitive.* The preferences are only evidence. Which of these health states is worse is not constituted by any of the preferences.

In addition to the above criticisms of evaluating health states by measuring preferences, it is worth recalling all the general difficulties with identifying well-being with the satisfaction of preferences also apply. Enough has already been said about conflict between well-being and preference satisfaction, when preferences are based on false beliefs or when preferences are in some sense distorted. But there are plenty of other difficulties. For example, to regard welfare as the satisfaction of preferences leads to complications when preferences change. As Richard Brandt (1979, ch. 13) and Derek Parfit (1984, ch. 8) point out, if an individual's preference ranking changes, then it is unclear whether the individual is made better off by satisfying or frustrating the original preferences. This theoretical problem is linked to a practical problem in circumstances in which policies and institutions have systematic effects on preference rankings. Assessments of health policies would then have to depend in part on one's views concerning *which* preferences to promote or concerning which institutions provide a suitable framework within which desirable preferences will develop (Michael McPherson 1982, 1983). Should one be concerned about the extent to which one satisfies current preferences, when one judges that they are in any event likely to change? Should one aim to modify preferences to make them easier to satisfy? A preference satisfaction theory of well-being makes some of these questions very difficult to answer, and it suggests implausible answers to others. In particular, a preference-satisfaction view of well-being apparently has the alarming implication that (other things being equal) we should educate people to have easily satisfied preferences. It implies implausibly that those with low expectations of health will tend to be "healthier," which is indeed what one sometimes finds if one takes literally the results of polls (Murray 1996, p. 25).

Before ending this section, let me make one last observation. The argument that evaluative questions cannot be settled by measuring preferences or subjective states is equally valid and much more obvious, when one shifts from questions concerning the evaluation of health states – where preferences and feelings are relevant evidence – to questions concerning the allocation of health care, the distribution of the burdens of paying for health care, the rationing of scarce organs among those in need of a transplant, and so forth. However such moral questions may be understood, they are not questions about preferences or mental states and cannot be answered by measuring preferences or by measuring mental states. I have not, however, addressed the issue yet of whether

these questions, along with questions concerning the evaluation of health states, should be answered by eliciting the opinions of those concerned; and it is to that issue that I turn in the next two sections.

6 Evaluative questions and sociological questions

Thus far I have criticized the proposal that health economists can replace the moral conundrums involved in evaluating health states with the measurement of the preferences of members of the target population. We have seen that one cannot evaluate health states by measuring preferences, because well-being is not the satisfaction of preferences and because the evaluation of health states does not depend exclusively on well-being. The matter is not settled by preferences. Judgment is needed. If “asking” is interpreted as measuring *preferences*, then one cannot evaluate health states or answer moral questions in general by just asking.

As I have just argued, attempts to “measure preferences” by the use of rating scales, time trade-offs, standard gambles, and person trade offs should in many cases be regarded as attempts to elicit opinions. The rationale for “asking” cannot then be that one is determining the (true) severity of health states by measuring well-being (which consists in the satisfaction of preference). If health economists base their evaluation of health states on the opinions they elicit when they purport to measure preferences, then they have abandoned the attempt to specify the true severity of health states. (Perhaps some would even deny that there is any such thing). Instead of seeking the answers to ethical questions in measuring preferences, health economists would appear to be *replacing* the ethical questions with sociological questions, which they seek to answer by eliciting opinions. Although Erik Nord speaks of measuring preferences, he is not concerned with preferences for health states. He is instead concerned to find out what members of the target population want of their health care system and, in light of those principles, how they would judge different allocations of health-care resources.

Recall that, unlike the WHO, Nord is not concerned to measure health. He is concerned with how to set priorities within the health system. The reason for “asking” – for eliciting opinions – is that those priorities should conform to the wishes of those who are served by the health system. For example, whether any of the health-care budget should go to fund expensive procedures such as transplants depends on what people want and on what they believe to be just, not on whether doing so produces a higher or lower level of population health. Just as members of a wine-tasting society should be provided with red wine if that is what they chose to pool their money to purchase, so members of some society or insurance group should be provided with liver transplants if that is what

they chose to pool their money to pay for. So health policy needs “empirical ethics” – the attempt to resolve moral questions by finding out what the values of the target population is. Notice that this is supposed to apply to all the many evaluative questions that arise concerning health, not just to the evaluation of health states and interventions.

Empirical ethics might be understood in two different ways. On one interpretation, it is the conventionalist or subjectivist view that the answers to moral questions are settled by the beliefs of the target population. If nearly everybody in the population thinks that it is worse to be blind than to be deaf, then it is on average worse in that population to be blind than to be deaf. On a second interpretation, empirical ethics says nothing about how to find the correct answers to moral questions. Instead it addresses the question of how decisions should be made. Whether or not it is worse to die of heart failure or of lung cancer, the distribution of resources to the treatment of those conditions should be determined democratically. The lives of members of the target population are at stake, and they are paying. So their values, whether right or wrong, should govern. In this section, I shall address the first interpretation of empirical ethics, reserving the discussion of the second interpretation to the next section.

Although the term, “empirical ethics” is employed by Jeff Richardson (2000a, 2000b), neither he nor Nord defend the extreme views discussed in this section and the next. Indeed Richardson argues explicitly that people’s judgments are sometimes mistaken and that they should sometimes be overruled (Richardson 2000a). But the extreme view is tempting and thus worth presenting and criticizing. I should emphasize that those who – like me – reject empirical ethics – who deny that ethical questions can be answered by eliciting opinions – do not deny that health systems should be responsive to the concerns of the populations they serve. There is no controversy about the need to find out what people want of their health systems, and it could be that Nord and Richardson are mainly concerned to emphasize this need. Whether or not they mean anything more, I shall not identify empirical ethics with merely finding out what people want of their health systems. I take empirical ethics instead to be the controversial view that moral questions should be resolved empirically.

Having thus clarified what is at issue, here are five arguments against the view that moral questions can be answered by determining what people’s moral judgments are.

First, empirical ethics in this sense sharply conflicts with the settled convictions of most people. Social approval of slavery, female infanticide, female genital mutilation, or ethnic cleansing does not make any of these practices right. Empirical ethics conflicts with the settled convictions of most people, those who think that social consensus should be decisive need to explain why their view has not thereby undercut itself.

Second, empirical ethics places an unjustified weight on how one draws the boundaries of societies, groups, or populations. If the population is defined as the village, one answer may be correct. If it is the region, that answer may be wrong and a second answer correct. If it is the tribe, yet a third answer may be correct. Without principles for specifying the relevant group, the answers to moral questions become simply arbitrary. Furthermore, when there is no consensus, how are questions to be decided? By a simple majority vote? Should different intensities of feeling or preference matter? How can such questions concerning the empirical investigation of morals be answered without making the answers to moral questions arbitrary?

Third, when there is no consensus, how would the relativist decide moral questions? What role can *reasons* play within empirical ethics? If empirical ethics maintains that it is possible to give good reasons criticizing or defending ethical statements, then it must admit that a social consensus and the reasons that support it can both be mistaken. If, instead, empirical ethics denies that reasons have any role and holds that moral views are entirely matters of social convention, does it follow that any convention is just as good as any other? If so, it is difficult to explain what people are disagreeing about when they argue about ethics. If you and I disagree on some moral question, surely we are not arguing about what the majority of others think, and our disagreement cannot be settled by the results of a poll. If, on the other hand, not all conventions are equal and what is right depends on what conventions would be *best*, then one can make sense of moral disagreement and of the possibility of arguing for moral positions – but one will have undercut the view that moral questions can be answered through measuring attitudes.

Fourth, empirical ethics is inconsistent with the very nature of moral judgments, which are supposed to be rationally contestable, because it implies that the social consensus is always right, and minority views and the views of social reformers are always automatically mistaken. This means that one does not need to attend to the reasons or arguments minorities may give, since their positions are, merely in virtue of being minority views, automatically wrong – or at least until the moment when they become majority views and hence correct. Such a view clashes with the actual practice of moral argument and seems to leave no room for rational contestation of moral disagreements.

Fifth, empirical ethics also makes little sociological sense. A social consensus on moral issues is not just a “given” to be measured, rather than to be queried. Which moral views are accepted today depends on past *arguments* of moral reformers -- often as part of religious and political movements. It is a matter of sociological *fact* that moral views depend on reasons. It is thus mistaken to suppose that today’s consensus is no longer dependent on arguments. And if the consensus is subject to criticism, it cannot be automatically correct.

Philosophical reflection should not be regarded as opposed to social determinations of what is

right and wrong. Philosophical reflection instead plays a part *within* the social determination of what is right and wrong. For example, the consensus that once obtained in some parts of the world in favor of female circumcision is crumbling in part because of rational criticism. Philosophical reflection and argumentation are a part of the process whereby groups of people resolve -- and constantly reassess -- moral issues. If the values within a society are genuinely moral values, then they are sensitive to rational argument and criticism, and nothing prevents the analyst from making those arguments and criticisms. I am not arguing that the analyst or philosopher has any authority beyond that conveyed by his or her arguments. But neither is the analyst or philosopher required to bow to the social consensus. Those addressing ethical questions concerning health policy should defer to good argument, not to mere consensus.

It is worth underlining this general point with a specific illustration. Although I have been unable to get hold of transcripts of any groups deliberating concerning the evaluation of health states, Paul Dolan was kind enough to send me transcripts of four groups who were addressing the problem of allocating kidneys among those in need of transplants. This is obviously a difficult moral question. The transcripts involve groups of six respondents whose ages ranged from about 20 to 70. In their first meeting, groups discussed the general question of what factors ought to be taken into account in determining priorities for kidney transplants. At a second meeting, groups were given a much more specific problem, summarized in the table below. The first column says how long the patient will live while undergoing dialysis (and thus at a diminished quality of life). The second column says how long the patient will live with the transplant in a state of virtually full health. The third column states how much extra life the transplant provides. The fourth column, which was revealed to group members only midway through the discussion, gives the patient's age. After discussing the problem for a while knowing the ages, group members were asked to suppose that the six patients A - F were in fact the members of the group. Members of the group were told -- and reminded -- that each individual has only this one chance for a kidney. So one could not decide to give the kidney to F on the grounds that later some other kidney will become available for the remaining individuals later.

Name	No treatment	With treatment	Extra years	Age	Identity
A	16	46	30	20	
B	13	38	25	25	
C	10	30	20	35	
D	7	22	15	45	

E	4	14	10	50	
F	1	6	5	70	

The group members were asked to imagine that they constituted an advisory committee that would determine who got the transplant and who did not, and they took their hypothetical task seriously. Initially people in each group favored giving the kidney to F on the grounds that F would die soon without the transplant. But after further discussion and reminders from the facilitator that there would be no other kidney later on for A-E, most people opted for A on the grounds that the kidney does more good if it goes to A than if it goes to anyone else.

The groups probed many other issues. What caused the kidney failure? Should those whose kidney failed because of consumption of alcohol or drugs get a lower priority? (And should they get the lower priority on the grounds of their responsibility for their condition or on the grounds that they might destroy their new kidneys, too?) Once the ages were revealed, group members mentioned the consideration that F had already had a full life, while the younger patients had had comparatively little healthy life. Of special importance to the groups was the question of whether A-F had any dependents – mainly dependent children – and once A-F were identified with members of the groups, the consensus was to give the kidney to B or C rather than A, when B or C had young children.

The decisions of the groups relied on general principles, though these were rarely enunciated explicitly. Many of these are implicit in the last two paragraphs, and it is obvious that they may conflict. By making a decision, members of the groups suggested how they would resolve these conflicts. But there was no systematic probing of the priorities. Some of the discussion involves argument. Other parts of the discussion involve calling attention to features of the circumstances of different individuals. The facilitators had an important role both in explaining the stylized facts and in suggesting issues to discuss.

Although the groups reached similar conclusions, there were disquieting features to the discussions. Considerations that were relevant to some groups never even occurred to others. In response to the case for giving the kidney to B or C rather than A on the ground that B and C have children and A does not, the question arose in some groups, but not others, of the effect of A's condition on A's parents. Some groups raised the question of whether the profession or expected social contribution of patients ought to matter, while the issue was never mentioned in others. In discussing the relevance of dependents, a member of one group talked about his role in taking care of a dependent wife. In other groups, "dependent" was simply equated with "child."

Although the facilitators did sometimes raise issues that groups overlooked, they did not challenge the groups with counterarguments or with a demand to know the principles underlying the

decisions. In one of the discussions, when respondents were uncomfortable about having raised the question of whether the patient's profession should be relevant, the facilitator commented, "I mean there are no right or wrong answers, we're just here to find out what you think so, y'know, anything you think of, even if you afterwards say, "Well actually no, I don't want to take that into account." This could mean, "At this stage of the discussion, let's hear any ideas people might have concerning relevant factors and leave the assessment of those ideas until later," or it could mean "We are just trying to find out individual sentiments, which are not the sort of thing that could be right or wrong." In any event – as was true of many of the issues – discussion simply wandered away from the question of whether a person's profession should affect their chances of getting a transplant without ever reaching any conclusions.

In these transcripts one finds concrete moral philosophy mixed with anecdote and nervous meanderings. The group members are not philosophizing in the same way that philosophers or theologians would, but philosophers or theologians or others who have devoted themselves to moral reflection could play a useful part in the deliberations. Their concerns and arguments are continuous with those of the group members. It is a mistake to regard deliberative groups as a technique for measuring preferences that evades moral questions. Such groups instead attempt to answer moral questions in much the same way as do those who devote themselves to moral reflection. Philosophizing is a *part* of such groups, not something to be set *against* them.¹⁸

Empirical ethics is thus very much a counsel of despair. Moral questions concerning health and health policy, including the questions concerning the evaluation of health states upon which this essay has focused, are terribly difficult. Since they are so hard, it is tempting to turn from reflection on what is good and what is right to eliciting what members of some group think. But if health administrators, economists, philosophers, and theologians are all baffled, surely members of the target group who take the issues seriously must be baffled, too. Those who participate in deliberative groups muddle through somehow, but to suppose that all that is needed are mislabeled "preference measurements" would be irresponsible.

Because most adults possess a very considerable competency in the morality that governs their society, their moral judgments are raw material for philosophical articulation and criticism in much the same way that the judgments of grammaticality of native speakers are raw material for theories of grammar (Rawls 1971, pp. 49). Rather than beginning with the evaluations of indicator health states that result from some study involving time trade offs, standard gambles, person trade-offs or

18 Although more intricate and subtle than the discussion of deliberative groups, Francis Kamm's work in this volume (2001) shows with special clarity the continuity between "philosophical" and lay considerations of moral issues.

whatever, correcting for measurement error, and then proceeding to calculate the value of all health states, one needs to examine the *reasoning* involved in the evaluations, to identify the implicit principles, to assess them by considering their consequences and their consistency with other moral principles, to reveal specific judgments that are not consistent with the implicit principles, to assess both judgments and principles in terms of higher order principles concerning the nature of morality and the possibilities of gaining moral knowledge, and to make adjustments in all these (and hence in the evaluations of health states) until a genuine reflective equilibrium is reached.

In this process of reflective equilibrium, philosophical reflection plays a central role. One has abandoned empirical ethics and the vain hope that one could resolve moral questions concerning health and health policy by measurement rather than argument. Important roles for measurement remain to provide input into the process of reflective equilibrium and to provide information concerning what people want from their health systems. But one must abandon the relativist view that social consensus resolves moral questions, and one must face the ethical questions that economists and health administrators would be happy to avoid.

Before closing this section, it is worth emphasizing that rejecting empirical ethics and moral relativism does not imply arrogance, dogmatism, intolerance, or cultural imperialism. People who welcome criticism of their point of view, who are aware of their fallibility, and who are willing to change their minds when confronted with good arguments are not dogmatic. Those who are willing to look long and hard at the practices of other societies, who recognize that there are many different kinds of good lives, and who are particularly hesitant to coerce those who disagree are not intolerant or imperialist. Tolerance is a moral virtue – though it is, of course, not the most important of all virtues. It is not the same thing as skepticism or indifference. Someone who is tolerant in an intolerant society tries to undermine that intolerance. Unlike those committed to empirical ethics, who would have to conclude that intolerance is right wherever it is popular, those who value tolerance should fight for it, even in the face of an intolerant consensus.

Within the limits of what is morally permissible, health systems should conform to the wishes of those they serve. Consequently, those wishes must be investigated. In addition, the specific moral judgments that members of the target population are inclined to make, the sort of moral reasons they rely on, and the principles that are implicit in their judgments are all crucial raw materials in developing defensible answers to moral questions concerning health policy. So empirical investigation of the attitudes of the target population remains essential. Furthermore, measurement of preferences and of mental states provides relevant evidence concerning the character and consequences of health states. But neither measurement of preferences and feelings, nor elicitation of opinions settles evaluative questions. Policy analysts need to investigate how people understand the moral issues and

what reasons support their particular judgments. At the same time, policy analysts must also remind themselves that they may be misinformed, biased, mistaken, and blind to the varieties of good lives.

7 When and how to ask

Even if one grants the critique of empirical ethics made in the last section and recognizes that the social consensus may be wrong, one might still argue that the prevalent values should guide health policy. Just as many favor democracy, while recognizing that democratic procedures may lead to political mistakes, so one might insist that the values of the target population should govern health policy and the evaluation of health states, while recognizing that mistakes may result. Of course this argument for democracy has limits, since even the most democratic societies place some issues beyond the reach of majority vote.

The main reasons for arguing that certain decisions should be made democratically are the risks and disadvantages attached to any alternative and the principle that all competent adults should have an equal voice in deciding certain matters either as a matter of fundamental rights or because giving some a greater voice than others would be disrespectful toward those with a lesser voice. It would be treating them like children. According to the first reason, it is too dangerous to grant ultimate authority on policy matters to any group smaller than the whole of the citizenry, because that group could then dominate the society, abridge liberties, and endanger individual rights. According to the second view, regardless of the dangers of allowing decisions to be made undemocratically, doing so itself violates rights or fails to show respect. Notice that similar reasons underlie the limitations on what majorities can decide. Fundamental rights and liberties and a regime that secures the self-respect of citizens cannot be abrogated by majority vote. They are open to abrogation by super-majorities not because they are properly regarded as matters to be decided by the popular will, but because within a regime involving popular sovereignty no further protection is possible.

The arguments for democracy and for the limits to democracy apply to many health-care decisions. For example, it is, in my personal view, morally wrong when a rich society such as the United States fails to secure health insurance for all its citizens. But I think that decisions such as this one are properly made by the United States Congress, and that those who – like me – think that such decisions have been made mistakenly should do what they can *within* the political institutions of the United States to change those decisions.

Political decisions such as this one are, however, not at issue in defining a summary measure of population health or any of the other indices the WHO is working on. Judgments concerning the severity of people's health states are no more political choices than are judgments concerning their

real income. A concern for democracy dictates that decisions about what to *do* about health states or income distribution should be made democratically. It does not dictate that decisions about whether *A* is healthier than *B* or whether *A* is richer than *B* should be put up for a vote. Although the evaluations of health states for the purpose of determining health care priorities, which is what most health economists have been concerned with, are more directly aimed at policy, they are supposed to be inputs or information for decision making, not a substitute for decision making.

For example, suppose that half the population is suffering from illness *X* and half from illness *Y*, and that the quality of life with *X* is the same as the quality of life with *Y*. Suppose also that *X* and *Y* are not correlated with race, sex, social class, wealth or any of the other main dimensions of inequality within societies. There are limited resources and health care administrators have a choice between two interventions. The first completely cures illness *X* and does nothing for *Y*. The second cures one-quarter of those suffering from each of the diseases. I personally believe that the intervention that cures half the population should be preferred to the one that cures only a quarter of the population. In my view, there is nothing more unfair about being the unlucky ones who happen to come down with illness *Y* instead of *X* than being the unlucky ones who do not benefit from the treatment that cures one quarter of the population. Yet in many societies a majority would favor the second intervention on the grounds that it gives everybody a chance to be cured. Valuing democracy, those who think the second intervention inferior to the first would acquiesce in the democratic decision, but they would not conclude that the decision was correct. They would instead argue against the majority and explain how people are misapplying admirable principles demanding fair treatment to everyone. The majority does not determine which is the better policy. It determines only which policy is adopted.

Although moral questions concerning what is right and wrong are very different from most scientific questions, there is no stronger case for deciding moral questions democratically than there is for deciding scientific questions democratically. The democracy argument applies only to political actions. It does not apply to moral questions, including questions about how to evaluate health states. Those who support empirical ethics might complain that this view disenfranchises the population. Of course it does, because it maintains that the answers to moral questions, like the answers to factual questions, cannot be found by taking a vote.

Even though there is no violation of democracy if health states are not evaluated by measuring attitudes, the WHO's evaluation of health states may face political constraints. If the WHO's summary measure of population health does not seem reasonable to members of the populations whose health the WHO seeks to measure, then the measure will not be accepted. Whether or not depression is "in fact" worse than dementia, a measure of these health states must seem reasonable to

the relevant populations. So, the argument concludes, evaluations of health states must be determined by empirical investigation rather than by philosophical argument.

There is something to this argument, though not enough to vindicate the second interpretation of empirical ethics. What is right about the argument is that, for political reasons, the WHO's valuations cannot conflict too flagrantly with too many settled convictions of the world's populations. This constraint is however weak, because few people have settled convictions concerning the precise evaluation of different health states and improvements. There is a very sizeable margin of uncertainty and reasonable disagreement and thus a large space for direct consideration of ethical questions. Furthermore, even in cases in which the evaluations given by the WHO conflict with settled convictions of specific populations (as is the case for example with some gender-related health questions), the WHO's position can still command respect. If the arguments in defense of the WHO's rankings are good ones, as they are in case of insisting on similar evaluations of similar health states whether they mainly affect men or women, the WHO should attempt to make the case for its values rather than caving in to its critics. Furthermore, the need to justify the evaluations across different populations undercuts the claims of empirical ethics, which maintains that evaluations should conform to the attitudes within specific target populations. It is also worth noting that there is no guarantee that the evaluations that result from deliberations by focus groups will be acceptable to the population either.¹⁹

Although the constraints of acceptability do not bind very tightly, this might appear to be just good luck. I do not think it is luck at all, because I believe that the combination of the complexities involved in evaluating health states and the commonalities in the world's ethical systems – whether in sophisticated philosophical or theological presentations or in everyday life – guarantee a good deal of space for serious independent ethical theorizing concerning the severity of different health states.

What if I were wrong about this? What if the WHO faced a choice between an incorrect but politically acceptable evaluation of health states and a more accurate but unacceptable one? Depending on the details, the WHO might have to rely on the politically acceptable health-state evaluations. But it would be shirking its duties to the people of the world if it did not make clear why those formulating the summary measure of population health believed that the measure was in error, how large it took those errors to be, and how it would measure population health if it could persuade people to accept its evaluations.

¹⁹ This is not merely an abstract possibility. Paul Dolan tells me that deliberative groups in the U.K. regularly reach the conclusion that more health care resources should go to the young than to the old. Yet such a view is politically unacceptable.

8 Conclusions

There are four overall themes in this long essay. First, philosophical reflection is unavoidable concerning the many ethical problems posed by the attempt to measure health, to appraise health systems, and to answer a myriad of questions about health policy. This philosophical reflection need not be done by professional philosophers, and indeed most of it in fact is done by others, especially health-care providers and administrators, religious leaders and social commentators. Ethical reflection is a part of popular thinking about such questions rather than something opposed to it.

Second, many of these ethical questions are very hard. The evaluation of health states is a bewildering task. It is not clear what health is and how it is related to well-being or the quality of life. It is not clear what counts as well-being. It is not clear how time counts, how to render diminished health commensurable with death, or how to weight the several different dimensions along which health states diminish well-being. These questions are no easier for the non-professionals whom economists query than they are for health-care administrators or philosophers.

Third, it is a mistake to construe the task of evaluating health states as a matter of measuring preferences. Economists tend to emphasize preferences because most mistakenly measure well-being in terms of the satisfaction of preferences. Although the usual method of measuring preferences in terms of willingness to pay is obviously unacceptable in the domain of health and has been for the most part abandoned by health economists, the view that evaluating health states requires the measurement of preferences persists. But since well-being is not the satisfaction of preference and since the evaluation of health states does not depend only on the well-being of people who occupy those health states, economists should recognize the limited role that information about preferences plays in evaluating health states. They should realize that they are often eliciting opinions rather than measuring preferences, and they should abandon the hope, which some may harbor, that the evaluative tasks can be carried out by measuring preferences.

Fourth, ethical questions concerning health, health systems, and health policy cannot be answered by eliciting the opinions of the target population. Within the constraints of morality, health systems ought to provide target populations with what they want. Furthermore, the values, interpretations, arguments and principles accepted by the target population are part of the difficult task of deliberating on health-related moral questions. In these two ways, the values and wishes of the target population are of great importance. But they do not answer ethical questions. Empirical ethics is sociology, not ethics; and it is not good sociology. It is not ethics, because it only tells economists and health analysts what people want and what people think is right or wrong, good or bad. It does not tell them what is in fact right or wrong, good or bad. It is bad sociology, because it does not ask

the right questions. Those who think seriously about health need information about the reasoning, principles, and understanding of the target population, not just their attitudes concerning health states.

I am a philosopher, not an economist, a health administrator, or a health-care provider; and I am not even a specialist on ethical issues on health care. Nevertheless I would like to end on a constructive note with a positive suggestion about how to make progress on ethical questions concerning health care in the light of the four general themes defended in this essay. Here are two suggestions. The first is that the actually existing complex and messy interplay of experimentation, surveying, statistical projection, philosophical criticism, political haggling and protest, and so forth is itself one unlovely and unsystematic way of implementing the ideas defended here. I am not suggesting that the real world is the best of all possible worlds. It isn't. All I am suggesting is that despite its many deficiencies and despite the misconceptions involved in so many of the contributions to the dialogue, the actual process illustrates how ethical questions can be tackled without reducing them to questions of preference or of sociology.

A more speculative suggestion concerns the evaluation of health states. What protocol should be used in lieu of the WHO's? I am in no position to go into detail – and this essay is already long enough, but here are some rough ideas:

(1) Health states have to be evaluated reflectively and deliberately. Deliberative groups or active interviews are essential.

(2) For reasons of feasibility, these groups will only be able to evaluate a relatively small number of “indicator” conditions; and the methods by which values are imputed to other health states require careful scrutiny (which I am in no position to discuss here).

(3) Though the deliberative groups should be small enough to permit high-quality discussion, they should not be limited to health-care professionals. They should include people with a wide range of experiences and expertise. Either as members or as facilitators, groups should include health professionals, economists, and philosophers or religious leaders. Group membership should cross cultures.

(4) Deliberative groups need to be provided with a great deal of information concerning health states, including information about how they are experienced and information concerning typical preferences among health states. (Polling and measuring preferences and feeling can be extremely valuable or even essential as an input into deliberation, but should not be regarded as a plausible alternative to it.)

(5) The evaluative *judgments* of members of deliberative groups should be invoked in a number of different ways. *If* the task is to measure health, PTO1 and PTO2 ought not to be used, since these

trade offs involve questions of justice. Person trade-offs might be used, but groups should be asked to compare changes in the health states of groups that result from natural causes, rather than changes that result from health-care interventions, which invoke moral questions concerning the legitimacy of policies. Crucial to the design of methods for studying attitudes is the recognition that one is concerned to elicit opinions and arguments for and against those opinions, not preferences.

(6) More discussion at a higher level of rigor is needed within these deliberative groups. The reasoning supporting evaluation of health states should be made explicit, and effort should be devoted to making explicit and to evaluating the implicit principles that are being evoked. The evaluations groups make should be published, and critical arguments should be welcomed from as wide a group of people as is possible. Groups should reconvene to consider those critical arguments.

(7) Reconsideration of evaluations in the light of ranking the indicator conditions – as in the 1996 WHO protocol – is an excellent idea, and should be made as deliberative and reflective as possible. *Reasons* for conflicts should be carefully examined. Furthermore, after evaluations of indicator conditions have been tentatively settled, deliberative groups should be confronted with the implications of the evaluations of indicator conditions for the evaluation of other health states; and if those implications are unacceptable, the indicator conditions should be reevaluated, or the way in which the values of other health states are imputed should be revised.

(8) The WHO should not hope to establish some set of acceptable evaluations once and for all. The evaluation process will be expensive and arduous, and it will be constantly subject to challenge and revision. Insofar as is possible, the methods of computing the values of health states that are not directly evaluated should be easy to comprehend and to use, so that as modifications occur in the evaluation of indicator conditions or in the understanding of how the values of other states can be inferred from the values of indicator conditions, it will be easy to recalculate the values of health states generally.

Although this proposal is vague and speculative, it is not utopian. To the contrary, it is probably disheartening, because it makes clear how arduous, tedious, tenuous, and unstable the process of evaluating health states is likely to be. Those who want to get to work measuring aspects of health and health systems around the world may be shaking their heads at me and muttering, “All this is easy for you to say. You don’t have to do all the hard work.” And they would, of course, be right. No doubt many shortcuts will have to be taken.

My concern is not, however, to defend this last suggestion concerning how to elicit evaluative judgments. My themes are the four just summarized: that philosophical reflection is essential to address evaluative questions concerning health, that these questions are very difficult, that measuring preferences will not answer them, and that they cannot be answered by determining what members of

the target population think.²⁰

²⁰ I am deeply grateful for comments and suggestions from many members of the Goodness and Fairness Group. Special thanks to Bob Goodin, Erik Nord, Armando Ochangco, and Dan Wikler. I am also indebted to my colleague, Robert Streiffer, for helpful criticisms. Some of the arguments in sections 1 and 2 of this essay are drawn from Hausman and McPherson (1996), and I continue to be indebted to Michael McPherson for much of my understanding of contemporary welfare economics.

References

- Aristotle. 1962. *Nicomachean Ethics*. Tr. Martin Ostwald. Indianapolis: Bobbs-Merrill.
- Bentham, Jeremy. 1988 *Principles of Morals and Legislation*. Buffalo: Prometheus Books.
- Boorse, Christopher. 1975. "On the Distinction between Disease and Illness." *Philosophy and Public Affairs* 5: 49-68.
- _____. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44: 542-73.
- Brandt, Richard. 1979. *A theory of the right and the good*. Oxford: Oxford University Press.
- Costa and Rovira. Unpublished. "Eliciting Social Preferences for Health System Benefits: The Willingness to Assign Approach."
- Daniels, Norman. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy* 76: 256-82.
- _____. 1985. *Just Health Care*. Cambridge: Cambridge University Press.
- Dolan, Paul. 1998. "The Measurement of Individual Utility and Social Welfare." *Journal of Health Economics* 17: 39-52.
- Elster, Jon. 1983. *Sour grapes: studies in the subversion of rationality*. Cambridge: Cambridge University Press.
- Gauthier, David. 1986. *Morals by Agreement*. Pittsburgh: Pittsburgh University Press.
- Goodin, Robert. 1986. "Laundering Preferences." In Jon Elster and Anand Hylland, eds. *Foundations of Social Choice theory*. Cambridge, Cambridge University Press, pp. 75-101.
- T.H. Greene. 1969. *Prolegomena to Ethics*. Ed. A.C. Bradley. New York: Thomas Crowell Company.
- Griffin, James. 1986. *Well-being: its meaning, measurement and moral importance*. Oxford: Clarendon Press.
- Harris, John. 1987. "QALYfying the Value of Life." *Journal of Medical Ethics* 13: 117-23.
- Hausman, Daniel. 2000. "Revealed Preference, Belief and Game Theory." *Economics and Philosophy* 16: 99-115.
- Hausman, Daniel and Michael McPherson. 1996. *Economic Analysis and Moral Philosophy*. Cambridge: Cambridge University Press.
- Hicks, John. 1939. "The Foundations of Welfare Economics." *Economic Journal* 49: 696-712.
- Kaldor, Nicholas. 1939. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility." *Economic Journal* 49: 549-52.
- Kaplan, Robert and John Anderson. 1988. "A General Health Policy Model: Update and Applications." *Health Services Research* 23: 203-35.
- LeGrand, Julian. 1991. *Equity and Choice: An Essay in Economics and Applied Philosophy*. London:

Harper-Collins.

- Loomes, Graham and Lynda McKenzie. 1989. "The Use of QALYs in Health Care Decision Making." *Social Science and Medicine*. 28: 299-308.
- McPherson, Michael. 1982. "Mill's Moral Theory and the Problem of Preference Change." *Ethics* 92: 252-273.
- _____. 1983. "Want Formation, Morality, and Some 'Interpretive' Aspects of Economic Inquiry," in Norma Haan, Robert N. Bellah, Paul Rabinow, and William M. Sullivan, eds., *Social science as moral inquiry*. New York: Columbia University Press, pp. 96-124.
- Menzel, Paul, Marthe Gold, Jeff Richardson, Erik Nord and Jose-Louis Pinto-Prades. 1999. "Toward a Broader View of Values in Cost-effectiveness Analysis of Health." *The Hastings Center Report*.
- Mill, John Stuart. 1863. *Utilitarianism*. Rpt. in Marshall Cohen, ed. *The Philosophy of John Stuart Mill*. New York: Modern Library, 1961, pp. 321-98.
- Murray, Christopher. 1996. "Rethinking Dalys". In Christopher Murray and Alan Lopez. *The Global Burden of Disease*. Geneva and Cambridge, MA: WHO/Harvard University Press, pp. 1-98.
- Murray, Christopher and Alan Lopez. 1999. "Progress and Directions in Refining the Global Burden of Disease Approach." GPE Discussion Paper #1. Geneva: World Health Organization.
- Nord, Erik. 1994. "Seminarserie om veiledende verditall for prioritering ihelsevesenet (Workshops on a value table for prioritising in health care). Working Paper no. 1. Oslo: National Institute of Public Health.
- _____. 1999. *Cost-Value Analysis in Health Care: Making Sense out of QALYs*. Cambridge: Cambridge University Press.
- Nord, Erik, Jose Luis Pinto, Jeff Richardson, Paul Menzel and Peter Ubel. 1999. "Incorporating Societal Concern for Fairness in Numerical Valuations of Health Programmes." *Health Economics* 8: 25-39.
- Nozick, Robert. 1974. *Anarchy, State and Utopia*. Cambridge, MA: Harvard University Press.
- Olsen, Jan Abel. 1997. "Aiding Priority Setting in Health Care: Is There a Role for the Contingent Valuation Method?" *Health Economics* 6: 603-12.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Patrick, D., J. Bush, and M. Chen. 1973. "Methods for Measuring Levels of Well-being for a Health State Index." *Health Services Research* 8: 228-45.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Robbins, Lionel. 1935. *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.
- Rosser, R. and P. Kind 1978. "A Scale of Valuations of State of Illness: Is There a Social Consensus?"

International Journal of Epidemiology 7: 347-58.

Sen, Amartya. 1971. "Choice functions and revealed preference." *Review of Economic Studies* 38: 307-17.

_____. 1973. "Behaviour and the concept of preference." *Economica* 40: 241-59.

Sidgwick, Henry. 1901. *The methods of ethics*. 6th ed. London: Macmillan.

Summers, Lawrence. 1992. "Memorandum," *The Economist*. February 26.