

## 2

## UTILITARIANISM

It is generally accepted that the recent rebirth of normative political philosophy began with the publication of John Rawls's *A Theory of Justice* in 1971, and his theory would be a natural place to begin a survey of contemporary theories of justice. His theory dominates contemporary debates, not because everyone accepts it, but because alternative views are often presented as responses to it. But just as these alternative views are best understood in terms of their relationship to Rawls, so understanding Rawls requires understanding the theory to which he was responding—namely, utilitarianism. Rawls believes, rightly I think, that in our society utilitarianism operates as a kind of tacit background against which other theories have to assert and defend themselves. So that is where I too will begin.

Utilitarianism, in its simplest formulation, claims that the morally right act or policy is that which produces the greatest happiness for the members of society. While this is sometimes offered as a comprehensive moral theory, I will focus on utilitarianism as a specifically political morality. On this view, utilitarian principles apply to what Rawls calls 'the basic structure' of society, not to the personal conduct of individuals. However, since much of the attraction of utilitarianism as a political morality stems from the belief that it is the only coherent and systematic moral philosophy, I will briefly discuss some features of comprehensive utilitarianism in section 3. In either its narrow or comprehensive version, utilitarianism has both devoted adherents and fierce opponents. Those who reject it say that the flaws of utilitarianism are so numerous that it cannot help but disappear from the landscape (e.g. Williams 1973). But there are others who find it hard to understand what else morality could be about than maximizing human happiness (e.g. Hare 1984).

## 1. TWO ATTRACTIONS

I will start with utilitarianism's attractions. There are two features of utilitarianism that make it an attractive theory of political morality. First, the goal

which utilitarians seek to promote does not depend on the existence of God, or a soul, or any other dubious metaphysical entity. Some moral theories say that what matters is the condition of one's soul, or that one should live according to God's Divine Will, or that one's life goes best by having everlasting life in another realm of being. Many people have thought that morality is incoherent without these religious notions. Without God, all we are left with is a set of rules—'do this', 'do not do that'—which lack any point or purpose.

It is not clear why anyone would think this of utilitarianism. The good it seeks to promote—happiness, or welfare, or well-being—is something that we all pursue in our own lives, and in the lives of those we love. Utilitarians just demand that the pursuit of human welfare or utility (I will be using these terms interchangeably) be done impartially, for everyone in society. Whether or not we are God's children, or have a soul, or free will, we can suffer or be happy, we can all be better or worse-off. No matter how secular we are, we cannot deny that happiness is valuable, since it is something we value in our own lives.

A distinct but related attraction is utilitarianism's 'consequentialism'. I will discuss what exactly that means later on, but for the moment its importance is that it requires that we check to see whether the act or policy in question actually does some identifiable good or not. We have all had to deal with people who say that something—homosexuality, for example (or gambling, dancing, drinking, swearing, etc.)—is morally wrong, and yet are incapable of pointing to any bad consequences that arise from it. Consequentialism prohibits such apparently arbitrary moral prohibitions. It demands of anyone who condemns something as morally wrong that they show *who is wronged*, i.e. they must show how someone's life is made worse off. Likewise, consequentialism says that something is morally good only if it makes someone's life better off. Many other moral theories, even those motivated by a concern for human welfare, seem to consist in a set of rules to be followed, whatever the consequences. But utilitarianism is not just another set of rules, another set of 'dos' and 'don'ts'. Utilitarianism provides a test to ensure that such rules serve some useful function.

Consequentialism is also attractive because it conforms to our intuitions about the difference between morality and other spheres. If someone calls certain kinds of consensual sexual activity morally wrong because they are 'improper', and yet cannot point to anyone who suffers from them, then we might respond that the idea of 'proper' behaviour being employed is not a *moral* one. Such claims about proper behaviour are more like aesthetic claims, or an appeal to etiquette or convention. Someone might say that punk rock is 'improper', not legitimate music at all. But that would be an aesthetic criticism, not a moral one. To say that homosexual sex is 'improper', without being able to point to any bad consequences, is like saying that Bob Dylan

sings improperly—it may be true, but it is not a moral criticism. There are standards of propriety that are not consequentialist, but we think that morality is more important than mere etiquette, and consequentialism helps account for that difference.

Consequentialism also seems to provide a straightforward method for resolving moral questions. Finding the morally right answer becomes a matter of measuring changes in human welfare, not of consulting spiritual leaders, or relying on obscure traditions. Utilitarianism, historically, was therefore quite progressive. It demanded that customs and authorities which had oppressed people for centuries be tested against the standard of human improvement ('man is the measure of all things'). At its best, utilitarianism is a strong weapon against prejudice and superstition, providing a standard and a procedure that challenge those who claim authority over us in the name of morality.

Utilitarianism's two attractions, then, are that it conforms to our intuition that human well-being matters, and to our intuition that moral rules must be tested for their consequences on human well-being. And if we accept those two points then utilitarianism seems to follow almost inevitably. If human welfare is the good which morality is concerned with, then surely the morally best act is the one which maximizes human welfare, giving equal weight to each person's welfare. Those who believe that utilitarianism has to be true are convinced that any theory which denies either of these two intuitions must be false.

I agree with the two core intuitions. If there is a way to challenge utilitarianism, it will not take the form of denying these intuitions. A successful challenge will have to show that some other theory does a better job of spelling them out. I will argue later that there are other theories which do just this. But first we need a closer look at what utilitarianism amounts to. Utilitarianism can be broken down into two parts:

1. an account of human welfare, or 'utility', and
2. an instruction to maximize utility, so defined, giving equal weight to each person's utility.

It is the second claim which is the distinctive feature of utilitarianism, and it can be combined with various answers to the first question. So our final judgement of utilitarianism will depend on our evaluation of the second claim. But it is necessary to begin by considering various answers to the first question.

## 2. DEFINING UTILITY

How should we define human welfare or utility? Utilitarians have traditionally defined utility in terms of happiness—hence the common but misleading slogan 'the greatest happiness of the greatest number'.<sup>1</sup> But not every utilitarian has accepted such a 'hedonistic' account of human welfare. In fact, there are at least four identifiable positions taken on this question.

### (a) Welfare hedonism

The first view, and perhaps the most influential in the utilitarian tradition, is the view that the experience or sensation of pleasure is the chief human good. It is the one good which is an end-in-itself, to which all other goods are means. Bentham, one of the founders of utilitarianism, said, in a famous quote, that 'pushpin is as good as poetry' if it gives the same intensity and duration of pleasure. If we prefer poetry to pushpin, if we think it a more valuable thing to do with our time, it must be because it gives us more pleasure.

This is a dubious account of why we prefer some activities over others. It is a cliché, but perhaps a true one, that poets often find writing to be painful and frustrating, yet they think it is valuable. This goes for reading poetry as well—we often find poetry disturbing rather than pleasurable. Bentham might respond that the writer's happiness, like the masochist's, lies precisely in these apparently unpleasant sensations. Perhaps the poet really finds pleasure in being tortured and frustrated.

I doubt it. But we do not have to settle that question, for Robert Nozick has developed an even stronger argument against welfare hedonism (Nozick 1974: 42–5; cf. Smart 1973: 18–21). He asks us to imagine that neuropsychologists can hook us up to a machine which injects drugs into us. These drugs create the most pleasurable conscious states imaginable. Now if pleasure were our greatest good, then we would all volunteer to be hooked for life to this machine, perpetually drugged, feeling nothing but happiness. But surely very few people would volunteer. Far from being the best life we can lead, it hardly counts as leading a life at all. Far from being the life most worth leading, many people would say that it is a wasted life, devoid of value.

In fact, some people would prefer to be dead than to have that sort of life. Many people in the United States sign 'living wills' which demand that they be taken off life support systems if there is no hope of recovery, even if those systems can remove pain and induce pleasure. Whether or not we would be better off dead, we would surely be better off undrugged, doing the things we think worth doing in life. And while we hope we will be happy in doing them, we would not give them up, even for guaranteed happiness.

### (b) Non-hedonistic mental-state utility

The hedonistic account of utility is wrong, for the things worth doing and having in life are not all reducible to one mental state like happiness. One response is to say that many different kinds of experiences are valuable, and that we should promote the entire range of valuable mental states. Utilitarians who adopt this account accept that the experience of writing poetry, the mental state accompanying it, can be rewarding without being pleasurable. Utilitarianism is concerned with all valuable experiences, whatever form they take.

But this does not avoid Nozick's objection. Nozick's invention is in fact called an 'experience machine', and the drugs can produce any mental state desired—the ecstasy of love, the sense of accomplishment from writing poetry, the sense of peace from religious contemplation, etc. Any of these experiences can be duplicated by the machine. Would we now volunteer to get hooked up? The answer is still, surely, no.

What we want in life is something more than, or other than, the acquisition of any kind of mental state, any kind of 'inner glow', enjoyable or otherwise. We do not just want the experience of writing poetry, we want to write poetry; we do not just want the experience of falling in love, we want to fall in love; we do not just want the feeling of accomplishing something, we want to accomplish something. It is true that when we fall in love, or accomplish something, we also want to experience it. And we hope that some of those experiences will be happy. But we would not give up the opportunity to fall in love, or accomplish something, even for the guaranteed experience of those things inside an experience machine (Lomasky 1987: 231–3; Larmore 1987: 48–9; Griffin 1986: 13–23; Finnis 1981: 85–8).

It is true that we sometimes just want certain experiences. That is one reason people take drugs. But our activities while undrugged are not just poor substitutes for getting what drugs can give us directly. No one would accept that mental states are all that matter, such that being hooked up to an experience machine would be the fulfilment of their every goal in life.

### (c) Preference satisfaction

Human well-being is something more than, or other than, getting the right sequence of mental states. A third option is the 'preference-satisfaction' account of utility. On this view, increasing people's utility means satisfying their preferences, whatever they are. People may want to experience writing poetry, a preference which can be satisfied in the experience machine. But they may also want to write poetry, and so forgo the machine. Utilitarians who adopt this account tell us to satisfy all kinds of preferences equally, for they equate welfare with the satisfaction of preferences.

However, if the first two views leave too much out of their account of well-being, this third view leaves too much in. Satisfying our preferences does not always contribute to our well-being. Suppose that we are ordering food for lunch, but some of us want pizza, while others want Chinese food. If the way to satisfy the most preferences is to order pizza, then this sort of utilitarianism tells us to order it. But what if, unbeknownst to us, the pizza we ordered is poisoned, or just rancid? Ordering it now would not promote our welfare. When we lack adequate information, or have made mistakes in calculating the costs and benefits of a particular action, then what is good for us can be different from the preferences we currently have.

Preferences, therefore, do not define our good. It is more accurate to say that our preferences are predictions about our good. We want to have those things which are worth having, and our current preferences reflect our current beliefs about what those worthwhile things are. But it is not always easy to tell what is worth having, and we could be wrong in our beliefs. We might act on a preference about what to buy or do, and then come to realize that it was not worth it. We often make these sort of mistakes, both in specific decisions, like what food to order, and in 'global preferences' about what sort of life to lead. Someone who has planned for years to be a lawyer may get to law school and realize that they have made a mistake. Perhaps they had a romantic view of the profession, ignoring the competitiveness and drudgery involved. Someone who had planned to remain in their hometown may come to realize that it is a parochial way to live, narrow and unchallenging. Such people may regret the years they spent preparing for a certain way of life, or leading that life. They regret what they have done, because people want to have or do the things which are *worth* having or doing, and this may be different from what they *currently prefer* to have or do. The first is what matters to us, not the second (Dworkin 1983: 24–30; 2000: 242–54).

Utilitarianism of the preference-satisfaction variety says that something is made valuable by the fact that lots of people desire it. But that is wrong, and indeed backwards. Having the preference does not make it valuable—on the contrary, its being valuable is a good reason for preferring it. And if it is not valuable, then satisfying my mistaken preference for it will not contribute to my well-being. My utility is increased, then, not by satisfying whatever preferences I have, but by satisfying those preferences which are not based on mistaken beliefs.

A related problem with the preference-satisfaction approach is the phenomenon of 'adaptive preferences', in which people who cannot achieve some desired goal gradually lose their desire for it. This is known as the 'sour grapes' problem, after Aesop's fable about the fox who, after repeated failed attempts to reach the grapes overhead, declares that he does not want them anyway since they are probably sour. It is difficult to live with the disappointment of

unsatisfied preferences, and one way to deal with this disappointment is to persuade oneself that the unattainable goal was not in fact worth seeking. The extreme version of this phenomenon is the case of the 'contented slave', who adapts to her enslavement by claiming she does not want freedom. There is some debate whether there really were such contented slaves, but the general phenomenon of adaptive preferences is well established in psychological and social science studies (Elster 1982*b*; 1983*a*). It also arises, for example, in accounts of attitudes towards traditional gender roles. The more difficult it is for people to imagine changing these roles, the more likely they will adapt their preferences so as to desire only those things which are consistent with these roles.<sup>2</sup>

This raises a serious problem for evaluating political institutions in terms of their ability to satisfy people's preferences. If people adapt their preferences to what they can realistically hope to achieve, then even a repressive society that denies important opportunities for fulfilment to large numbers of people may nonetheless do well in satisfying people's (adapted) preferences. In fact, it may do better than an open and democratic society which prides itself on giving freedom and opportunity to all citizens. It is quite possible that there are more unsatisfied preferences in a free society than in a repressive society that teaches people from birth not to desire certain things.

#### (d) Informed preferences

The fourth account of utility tries to accommodate the problem of mistaken and adaptive preferences by defining welfare as the satisfaction of 'rational' or 'informed' preferences. Utilitarianism, on this view, aims at satisfying those preferences which are based on full information and correct judgements, while filtering out those which are mistaken and irrational. We seek to provide those things which people have good reason to prefer, that really make their life better off.

This fourth account seems right—the chief human good is the satisfaction of rational preferences.<sup>3</sup> But while this view is unobjectionable, it is extremely vague and difficult to apply or measure. Happiness at least had the merit of being in principle measurable. We all have a rough idea of what would increase happiness, what would increase the ratio of pleasurable to painful sensations. A pleasure machine would do that best. But once we view utility in terms of satisfying informed preferences, we have little guidance.

For one thing, how do we know what preferences people would have if they were informed and rational? Which religious beliefs, for example, would informed people hold? How do we know when a desire to follow a traditional gender role is an authentic expression of the person's good, as opposed to a merely adaptive preference? What sort of 'time-discounting' is rational—i.e. is it irrational to care more about what happens to me today than about what

will happen to me tomorrow? The issues involved are complex, yet we need an answer in order to begin the utilitarian calculations.

Moreover, even if we know which preferences are rational, there are many different *kinds* of informed preferences, with no obvious way to aggregate them. How do we weigh career accomplishment against romantic love, if there is no single overarching value like happiness to measure them by? The two goods may be 'incommensurable'—not measurable on any single scale.<sup>4</sup>

More puzzling yet is the fact that we have dropped the 'experience requirement'—i.e. informed preferences can be satisfied, and hence our utility increased on this fourth account, without it ever affecting our conscious experiences. Richard Hare, for example, argues that my life goes worse if my spouse commits adultery, even if I never come to know of it. My life is made worse because something that I wanted not to happen has happened. This is a perfectly rational and informed preference, yet my conscious experience may not change whether it is satisfied or left unsatisfied (Hare 1971: 131).

I agree with Hare that 'unexperienced' preferences should count in determining well-being. It really does make my life worse when my preferences are violated without my knowing it. For example, if I continue to act towards my spouse on the belief that she has not committed adultery, then I am now acting on a falsehood. I am living a lie, and we do not want to live such a life (Raz 1986: 300–1). We often say of others that what they do not know will not hurt them. But it is hard to think that way of our own good. I do not want to go on thinking I am a good philosopher if I am not, or that I have a loving family if I do not. Someone who keeps the truth from me may spare me some uncomfortable conscious experiences, but the cost may be to undermine the whole point of my activities. I do philosophy because I think I do it well. If I am not doing it well, then I would rather do something else. I do not want to continue on the mistaken belief that I am doing it well, for I would be wasting my time, and living a lie, which are not things I want to do. If I were to discover that my belief is false, then my activity would have lost its point. And it would have lost its point, not when I discovered that the belief was not true, but when it ceased to be true. At that point, my life became worse off, for at that point I could no longer achieve the goals I was concerned to pursue.

Or consider the desires of parents regarding their children. As James Griffin notes, 'if a father wants his children to be happy, what he wants, what is valuable to him, is a state of the world, not a state of his mind; merely to delude him into thinking that his children flourish, therefore, does not give him what he values' (Griffin 1986: 13). His life is worse off if his children are suffering, even if he is blissfully unaware of this suffering.

We must accept the possibility that our lives can go worse even when our conscious experiences are unaffected. But this leads to some strange results. For example, Hare extends the notion of utility to include the preferences of

dead people. I may have a rational preference that my reputation not be labelled when I am dead, or that my body not be left to rot. It seems bizarre to include the preferences of dead people in utility calculations, but what distinguishes them from the preference that one's spouse not commit unknown adultery? In both cases, we have rational preferences for things which do not affect our conscious states. Not every action which goes against a dead person's preferences makes their life worse off, but where will we draw the line? And how can we weigh the preferences of the dead against the preferences of the living?<sup>5</sup>

In short, the 'informed preference' account is plausible in principle, but very difficult to apply in practice. There are difficulties both in determining which preferences increase welfare when satisfied (i.e. which preferences are 'rational' or 'informed'), and in measuring levels of welfare even when we do know which preferences are rational (i.e. comparing 'incommensurable' forms of utility). As a result, we may find ourselves in a situation where it is impossible to know which act maximizes utility, either for a given individual or for society at large.

Some people have concluded from this that utilitarianism must be rejected. If we accept the fourth view of welfare as the satisfaction of informed preferences, and if welfare cannot be clearly identified or aggregated on that view, then there is no way to know which act maximizes welfare, and we need some other account of the morally right act.

But this argument is, if anything, too strong. After all, these difficulties of identifying and balancing informed preferences arise not only in utilitarian moral reasoning, but in any form of prudential reasoning about how to lead our lives (Bailey 1997: 18–19). We constantly need to make decisions about how to balance different kinds of goods, over different time-frames, and to make judgements about how our life can go better or worse. If we have no rational basis for making these judgements, due to our lack of information or the incommensurability of goods, then it is the entire structure of prudential reasoning, not just utilitarianism, which is at risk. In reality, however, we do make these decisions, more or less successfully, even if we have no procedure for guaranteeing that our preferences are truly informed, and no mathematical formula for adding up all the different kinds of goods that are in our life.

To be sure, utilitarianism as a political philosophy requires that we be able to compare utility gains and losses across lives, not just within a particular life. In order to decide who should be given scarce resources, we may need to judge whether A's potential fulfilment outweighs B's disappointment. This is the problem of the 'interpersonal comparability' of utility, and some people think that, even if we can make rational judgements about how to maximize utility within a single life, we cannot do so across lives. We cannot get inside other

people's heads to know whether our fulfilments and disappointments are greater or lesser than theirs.<sup>6</sup>

But here again, this is too quick. If we were unable to make utility comparisons across lives, then we would be unable to make rational decisions about whether or when to help our friends, neighbours, or even our children. Yet parents continually make judgements about whether the benefits to one child outweigh the burdens or disappointments imposed on another child or the parents themselves. It would require an extreme form of solipsism to assume that we cannot make rational judgements comparing utility across lives.

Moreover, there are various indirect ways to overcome these difficulties. For example, the informed preference approach tells us to filter out those preferences which are adaptive or irrational. In practice, however, there is no realistic way for the government to make this determination directly; it would require vast amounts of information about each person's background, capacities, emotional make-up, and so on. Indeed few people would want the government to be collecting this sort of information about them. The government can, however, deal with the problem of irrational or adaptive preferences in a more indirect manner: not by examining specific preferences of individuals, but rather by trying to ensure the appropriate conditions for the *genesis* of those preferences. We may not be able to identify which specific preferences are distorted by false beliefs or adaptive preferences but we can examine the social and cultural conditions under which people form and revise their preferences, to make sure that people have access to appropriate information, and/or opportunities to test alternative ways of life, and/or protection from false or distorting images or propaganda. We deal with the problem of false or adaptive preferences, not by directly filtering them out, but by eliminating the background conditions which generate such preferences. As we will see in later chapters, particularly the chapters on communitarianism and feminism, many debates in contemporary political philosophy revolve precisely around these questions about the appropriate background conditions for the genesis of our preferences.

Similarly, there may be indirect ways of resolving the problem of interpersonal comparability. In theory, utilitarianism says that we should directly compare the welfare gains and losses of different people. In reality, this is impossible—the government cannot get inside the heads of citizens to weigh the relative strength of their joys and disappointments. However, for public policy purposes, we can adopt a more indirect strategy. We can ignore the details of individuals' preferences and focus instead on the all-purpose goods like liberties and resources which are useful to people whatever their more specific preferences. We can then use the distribution of these all-purpose goods as a reasonable proxy for the distribution of preference satisfaction (Goodin 1995: 13, 20–1). We measure gains and losses to individuals, not by

examining increases or decreases in their level of preference satisfaction, but by measuring increases or decreases in the level of all-purpose means they can use to satisfy their preferences.

Utilitarianism, on this view, would not aim at maximizing the satisfaction of people's preferences directly, but rather indirectly, by maximizing the overall amount of all-purpose goods available to people to satisfy their preferences. As we will see, this 'resourcist' solution to the problem of interpersonal comparability is adopted by most liberal theories of justice, and indeed is preferable not only on grounds of convenience and feasibility, but also on grounds of responsibility (see Ch. 3, pp. 72–4 below).

So the logistical problems confronting utilitarianism are serious, but not fatal. No doubt there will be some cases where we simply cannot determine which act maximizes utility, and hence cannot determine which act is morally right, on utilitarian principles. But as we will see, this is a problem that arises for most political theories. There is no reason to exclude the possibility that humans may not always be able to determine the morally right act. In any event, even if there is an inherent incommensurability of different kinds of value, such that we cannot say that one of a range of value-increasing acts maximizes value, we can still make some less fine-grained rankings, and so make judgements about better or worse acts (Griffin 1986: 75–92).

So utilitarianism, despite its traditional ties to welfare hedonism, is compatible with any of the four accounts of utility. Of course, utilitarianism loses one of its attractions when it leaves hedonism behind. Once we reject the simple accounts of welfare as happiness or preference satisfaction, there is no straightforward method for measuring utility. Utilitarianism does not provide a uniquely simple criterion or scientific method to determine what is right and wrong. But while utilitarianism has no advantage over other theories in measuring human welfare, neither is it disadvantaged. Every plausible political theory has to confront these difficult questions about the proper account of human welfare, and nothing prevents utilitarianism from adopting whatever account its critics favour.<sup>7</sup> If we are to reject utilitarianism, then, it will have to be because of the second part of the theory—i.e. the instruction that we should maximize utility, whichever definition of utility we finally adopt.

### 3. MAXIMIZING UTILITY

Assuming that we have agreed on an account of utility, should we accept the utilitarian commitment to maximizing utility? Is this the best interpretation of our intuitive commitment to 'consequentialism'? Consequentialism tells us to be concerned with promoting people's utility, and, ideally, we would satisfy all the informed preferences of all people. Unfortunately, that is impossible.

There are limited resources available to satisfy people's preferences. Moreover people's preferences may conflict. So whose preferences should we satisfy? Consequentialism tells us to be concerned with consequences for human welfare, but what if the promotion of one person's welfare conflicts with that of another? Consequentialism needs to be spelled out if we are to answer that question.

How does utilitarianism spell out the idea that we should promote people's utility? Utilitarians say that the right action is the one that maximizes utility—e.g. that satisfies as many informed preferences as possible. Some people's preferences will go unsatisfied, if their preferences conflict with what maximizes utility overall. That is unfortunate. But since winners necessarily outnumber the losers, there is no reason why the preferences of the losers should take precedence over the more numerous (or more intense) preferences of the winners. For the utilitarian, equal amounts of utility matter equally, regardless of whose utility it is. No one stands in a privileged position in the calculations, no one has a greater claim to benefit from an act than any other. Hence we should bring about consequences which satisfy the greatest number of (informed) preferences amongst people in the society. (This, of course, is the barest sketch of the utilitarian account of consequentialism—I discuss two ways to flesh it out in the next section.)

This commitment to examining the consequences for human well-being is one of the attractions of utilitarianism, as compared to theories which say that we should follow tradition or divine law regardless of the human consequences. But the particular kind of consequentialism in utilitarianism is, I think, unattractive. Where it is impossible to satisfy all preferences, our intuitions do not tell us that equal amounts of utility should always have the same weight. Utilitarianism provides an oversimplified account of our commitment to consequentialism.

Before exploring these issues, however, there are some important differences within utilitarianism that need to be laid out. I have just said that, as utilitarians, we should seek to satisfy the greatest number of preferences. But as I mentioned earlier, there are two different accounts within utilitarianism of who the relevant 'we' is—on one view, all of us are obliged to act according to utilitarian principles, even in our personal conduct (comprehensive moral utilitarianism); on the other view, it is the major social institutions which are specifically obliged to act according to utilitarian principles (political utilitarianism). There are also two different accounts of what it means to 'act according to utilitarian principles'. On one view, this means that the agent should decide how to act by consciously making utilitarian calculations, by trying to assess how different actions would affect the satisfaction of informed preferences (direct utilitarianism); on the other view, the idea of maximizing utility enters only indirectly (if at all) into the agent's decision-making.

Morally right actions are those that maximize utility, but agents are more likely to maximize utility by following non-utilitarian rules or habits than by following utilitarian reasoning (indirect utilitarianism).

These two distinctions can be combined to generate different versions of utilitarianism. Utilitarian principles can be applied more or less comprehensively, and more or less directly. Much of the recent work on utilitarianism has been concerned with exploring these variations, and it seems clear that each version will generate different results. However, I believe that all versions share the same fundamental flaw. I will argue that there is something inherently unattractive about the utilitarian commitment to maximizing utility, and that this flaw is not substantially affected by how (directly or indirectly) or where (comprehensively or to politics) that commitment is applied.<sup>8</sup>

I will begin by considering some problems with utilitarianism as a comprehensive decision-procedure. If we view utilitarianism in this way, then the morally responsible agent will be what David Brink calls a 'U-agent'—someone who decides how to spend her time and resources by calculating the effects on overall utility of the various actions available to her (Brink 1986: 425). This sort of utilitarianism has few contemporary defenders, and many utilitarians would agree with the criticisms I am about to make. But I start with utilitarianism as a comprehensive decision-procedure because it raises in a particularly clear form problems that are also present in the more indirect and political versions of utilitarianism (s. 5). Moreover, the issues raised in this section, concerning the proper scope of personal relationships, will reappear in later chapters.

Imagine then that we are U-agents, and that we can calculate which act produces the most utility.<sup>9</sup> Should we base our actions on these utilitarian calculations? There are two main objections to utilitarian decision-making—it excludes the special obligations we have to particular people, and it includes preferences which should not be counted. These two problems stem from the same basic flaw, but I will examine them separately.

#### (a) Special relationships

U-agents who base their actions on utilitarian calculations assume that each person stands in the same moral relationship to them. But this does not allow for the possibility that I could have special moral relationships to my friends, family, lenders, etc., that I could be under a greater obligation to them than to other possible beneficiaries of my actions. Our intuitions tell us that there are such special obligations, and that they should be fulfilled even if those to whom I am not especially obligated would benefit more.

Consider a loan. It is part of our everyday morality that people come to have differential entitlements in virtue of having loaned money in the past. If someone lends me \$10, then she is entitled to receive \$10 back from me, even if

someone else could make better use of the money. Utilitarian reasoning disregards such backward-looking entitlements, for it says that only forward-looking consequences matter. For the U-agent, the moral value of an act lies solely in its causal properties of producing desirable states of affairs. Hence what I ought to do is pull on the causal lever which will produce the maximal amount of utility for the system as a whole. In deciding how to spend my \$10, I must look at all the potential preference satisfactions of people (including myself) and determine which action will maximize them. It is of no interest to the U-agent, in and of itself, that one of those people loaned me the \$10, or that someone else performed some service for me on the understanding that she would receive the money. It may be that if the utilities work out in a certain way, I ought to repay the loan, or fulfil my contract. But the process of deciding what to do will go on in exactly the same way as if I had not borrowed or promised the money.

This is counter-intuitive, for most of us would say that the 'past circumstances or actions of people can create differential entitlements or differential deserts to things' (Nozick 1974: 155). The person who lent me \$10 has, by that very act, acquired an entitlement to the \$10 I am now considering spending, even if some other use of the money would maximize happiness. Does this conflict with our view that morality should be about consequences for human welfare? No, for in saying that I should repay the loan, I am simply saying that I have a greater obligation, at this point in time, to promote my lender's welfare than to help others. We should repay the loan, not because we do not care about the harms and benefits which arise from that act, but because one benefit in particular has special weight.

Unlike the hard-line non-consequentialist, we need not say that these entitlements are indefeasible by any calculation of overall social consequences. If repaying the loan would somehow lead to nuclear destruction, then we clearly ought not to repay the loan. But we can say that there is a duty to repay loans and fulfil contracts which has some independent weight, to be considered alongside the moral weight of overall social benefits. The existence of past entitlements on the part of particular people partially pre-empts, or constrains, the utilitarian quest to maximize the general good. Averting a disastrous drop in welfare is a good reason for using the money in a different way, but the mere fact that repaying the loan does not maximally increase welfare is not a good reason. Not to repay the loan simply because it does not maximally increase utility is to ignore the special nature of our obligation to the lender.

This is so firmly entrenched in our moral consciousness that many utilitarians have tried to give a utilitarian account of the weight we attach to promises. They point out the many by-products of breaking a promise. For example, while someone other than the lender may be able to make better use



of the money, the lender will feel resentment at being deprived of a promised benefit, a disutility so great that it outweighs the increased utility achieved by giving the money to someone else (Hare 1971: 134). But this gets things backwards. We do not feel that breaking promises is wrong because it produces feelings of resentment. Rather, cheating on promises produces feelings of resentment because it is wrong (cf. Williams 1973: 143). Another utilitarian tactic is to point out that promises create expectations which people depend on. Moreover, failing to repay the loan will jeopardize the lender's willingness to lend in the future, and thereby jeopardize a valuable social institution. So utilitarians respond by pointing out that repaying loans is more likely to maximize utility than one might initially think (Sartorius 1969: 79–80).

This may be true, but it does not solve the problem. It still implies, for example, that 'if you have employed a boy to mow your lawn and he has finished the job and asks for his pay, you should pay him what you promised only if you cannot find a better use for your money' (Sartorius 1969: 79). The U-agent's reasoning, while more complex than one might initially think, still fails to recognize any special relationship between employer and employee, or lender and borrower. Some utilitarians are prepared to accept this result. Rolf Sartorius, for example, says that if the usual factors do not ensure that payment maximizes utility, i.e. if the boy 'is not likely to publicize my breaking my promise to him too loudly, appears to have a reservoir of trust in mankind generally, and any sum I could give him really would do more good if contributed to UNICEF, then the conclusion on act-utilitarian grounds must be that I should give the money to UNICEF. But is this really absurd?' (Sartorius 1969: 80). Yes, this is absurd. What is absurd here is not necessarily the conclusion but the fact that the boy's having actually performed the job, or that I had actually promised him the money, never enters into the decision as such. Notice that the consequences Sartorius mentions would be exactly the same even if the boy hadn't actually mowed the lawn, but simply (falsely) believed that he had done so, or falsely believed that I had promised him the money. The fact that the boy actually mowed the lawn, or that I had promised him the money, does not matter to the U-agent because nothing we could do or say could ever put us in a special moral relationship such that my obligation to him is greater than my obligation to others. No matter what the boy has done or I have said, he can never have a greater claim on my actions than anyone else.

In our everyday view, the existence of a promise creates a special obligation between two people. The U-agent, however, treats promises and contracts, not as creating special moral ties to one person, but as simply adding new factors into the calculation of overall utility. The everyday view says that I should repay loans *regardless* of whether it maximizes utility. The U-agent says that I should repay the loan *because* it maximizes utility. The boy has no greater

claim on me than others, he just is likely to benefit more than they are, and so repayment is the best way to fulfil my utilitarian obligation.

But that is not what a promise is—'to make a promise is not merely to adopt an ingenious device for promoting the general well-being, it is to put oneself in a new relation to one person in particular, a relation which creates a specifically new *prima facie* duty to him, not reducible to the duty of promoting the general well-being of society' (Ross 1930: 38). For U-agents, everyone (including oneself) stands in exactly the same moral position—i.e. everyone is an equally deserving possible beneficiary of one's actions. But this is too flat a picture of the moral landscape, for some people 'may also stand to [one] in the relation of promisee to promisor, of creditor to debtor, of wife to husband, of child to parent, of friend to friend, of fellow countryman to fellow countryman, and the like, and each of these relations is the foundation of a *prima facie* duty' (Ross 1930: 19).

The problem here goes deeper than an inadequate account of promises. The U-agent cannot accommodate the importance of any of our commitments. We all have commitments—to family, political causes, work—which form the focal point of our lives and give some identity to our existence. But if I am to act as a U-agent, then in each of my decisions, my commitments must be simply added in with all the projects of other people, and be sacrificed when I can produce more utility by promoting someone else's projects. That may sound admirably unselfish. But it is in fact absurd. For it is impossible to be genuinely committed to something and yet be willing to sacrifice that commitment whenever something else happens to maximize utility. Utilitarian decision-making asks that I consider my projects and attachments as no more worthy of my help than anyone else's. It asks, in effect, that I be no more attached to my commitments than to other people's. But that is no different from saying that I should not really be attached to my projects at all. As Bernard Williams puts it,

if you are a person who whole-heartedly and genuinely possesses some of these admirable [projects, affections, and commitments], you cannot also be someone in whose thought and action the requirements of utilitarianism are unfailingly mirrored, nor could you wish to be such a person. . . . utilitarianism must reject or hopelessly dilute the value of these other dispositions, regressing to that picture of man which early utilitarianism frankly offered, in which he has, ideally, only private or otherwise sacrificable projects, together with the one moral disposition of utilitarian benevolence. (Williams 1981: 51, 53)

Utilitarianism is therefore often said to be 'alienating', in the sense that it forces us to distance ourselves from the commitments and projects that give meaning to our lives.<sup>10</sup>

Of course, our projects and commitments should respect the legitimate



commitments of others. But the way to do this is not to consider them as having an equal claim on my time and energy to that of my own projects. Such an attitude is psychologically impossible, and undesirable even if possible. A valuable human life, on just about anyone's account of it, is one filled with attachments that structure one's life, that give some direction to it. It is the prospect of subsequent achievement or progress in such a commitment that makes our current actions meaningful. As a U-agent, however, one's actions will be determined almost wholly independently of one's commitments. The U-agent's decisions will be 'a function of all the satisfactions which he can affect from where he is: and this means that the projects of others, to an indeterminately great extent, determine his decision' (Williams 1973: 115). The U-agent will have few choices about how to lead his life, few opportunities to act on considerations of the kind of person he is, or wants to become. He will thus have little room for the things we associate with the very idea of 'leading a life'. These will all be submerged beneath the question of which causal levers are optimistic.

If I am to lead my own life, there must be room in which I am free to form my own commitments, including the sorts of contracts and promises discussed above. The problem of not allowing people to create special obligations to others through promises is just one aspect of the broader problem of not allowing people to set and pursue their own goals. The problem in all of these cases is the U-agent's assumption that each person has an equal claim to benefit from all of his actions.

Does our intuition in favour of meaningful commitments violate the idea that morality concerns consequences? No, for our intuitive commitment to the general idea of consequentialism never included a commitment to the continuous impartial determination of our actions by the preferences of others, to the exclusion of special relationships and projects. This is simply too crude an interpretation of our belief in consequentialism.

### (b) Illegitimate preferences

A second problem with utilitarianism as a decision-procedure concerns its demand, not that each person be given equal weight in our decision-making, but that each source of utility (e.g. each kind of preference) be given equal weight. Consider racial discrimination in a mainly white society. A government health care policy might plan to build one hospital for every 100,000 people, regardless of their race. But a number of whites prefer that blacks do not have equal health care, and when the utility calculations are done, it turns out that utility is maximized by depriving blacks of an equal share of health care (or school facilities etc.). Or what if the very sight of known homosexuals deeply offends the heterosexual majority? Perhaps utility is maximized if openly homosexual people are publicly punished and thrown in jail. Or what

about an alcoholic on skid row who has no friends, is offensive to many people, and a nuisance to everyone, begging for money and cluttering up public parks? Perhaps utility would be maximized if we quietly took such people and killed them, so they would not be seen, and would not be a drain on social resources in jail.

Some of these preferences are of course uninformed, and so satisfying them would not actually yield any utility (assuming we have abandoned the crude hedonistic accounts of utility). But the desire to deny the rights of others is not always uninformed, and even on the best account of utility, the satisfaction of these preferences can be a genuine source of utility for some people. As Rawls puts it, such preferences are 'unreasonable' from the point of view of justice, but are not necessarily 'irrational', from the point of view of an individual's utility (Rawls 1980: 528–30). If this sort of utility is counted, it may lead to discrimination against unpopular minorities.

Our everyday morality tells us that such preferences are unfair, and should not be counted. That racists want a group of people mistreated is no reason at all to give that group less health care. The racists' desire is illegitimate, so whatever utility would come from satisfying that preference has no moral weight. Even where there is no direct prejudice, there may be unfair preferences which should not count. Someone may wish that blacks do not move into their neighbourhood, not because they actively dislike blacks—they may not care one way or the other—but because others dislike blacks, and so the property value of their home will decrease. Such a preference that blacks be excluded from a neighbourhood is not prejudiced in the same way a racist's is. But it is still an illegitimate preference, since it requires that something be wrongfully taken from blacks. In all these cases, utility is maximized by discriminatory treatment, but only as a result of preferences for benefits which are wrongfully taken from others. Preferences like that, preferences for what rightfully belongs to others, have little or no weight in our everyday moral view.

Utilitarians do not accept the claim that preferences for what 'rightfully' belongs to others are illegitimate. For the U-agent there is no standard of what 'rightfully' belongs to anyone prior to the calculation of utility. What is rightfully mine is whatever distribution maximizes utility, so utility-maximizing acts by definition cannot deprive me of my rightful share. But this violates an important component of our everyday morality. Our commitment to the idea of consequentialism does not include a commitment to the idea that each source of utility should have moral weight, that each kind of preference must be counted.

It seems, then, that the U-agent, in trying to maximize utility, is violating, rather than spelling out, our intuitive idea of consequentialism. Some people deny that utilitarian decision-making has these counter-intuitive results. They

admit that utilitarian reasoning seems to allow, or even require, acts which violate special relationships or basic rights, whenever such acts would maximize utility. But they claim that these acts would be disallowed if we shifted to a more sophisticated form of utilitarian decision-making. I have been assuming that U-agents apply the test of utility maximization to particular acts. But 'rule utilitarians' argue that we should apply the test of utility to rules, and then perform whichever act is endorsed by the best rules, even if another act might produce more utility. Social cooperation requires rule-following, so we should assess the consequences, not simply of acting in a particular way on this occasion, but of making it a rule that we act in that way.<sup>11</sup>

The issue for U-agents, then, is to determine which set of rules is utility-maximizing. Are we better off in utilitarian terms following a rule that instructs us to keep promises, maintain special relationships, and respect rights, or following a rule that subordinates these principles to calculations of utility? The latter, utilitarians argue, would paradoxically decrease utility. It would make social cooperation difficult, create fear and insecurity, and cheapen the value of human life and liberty (Goodin 1995: 22; Singer 1977). Moreover, people are likely to abuse the power to break promises or discriminate in the name of the public good (Bailey 1997). Everyone is worse off if we adopt a rule to break promises or discriminate against unpopular groups whenever we think it would maximize utility.

Some commentators argue that rule-utilitarianism collapses into act-utilitarianism, since we can describe rules in such a detailed and narrow way as to make them equivalent to acts (Lyons 1965: ch. 4; Hare 1963: 130–6). Others dispute this (Harsanyi 1977b). But even if the distinction is valid, it seems unduly optimistic to assume that utility-maximizing rules will always protect the rights of weak and unpopular minorities. As Williams puts it, the assurance that justice will prevail is 'a tribute to the decency and imagination of those utilitarians but not to their consistency or their utilitarianism' (Williams 1972: 103).

In any event, this response does not really answer the objection, for even if it gets the right answer, it does so for the wrong reasons. On the rule-utilitarian view, the wrong done in discriminating against a minority group is the increased fear caused to others by having a rule allowing discrimination. The wrong done in not paying the boy who mowed my lawn is the increased doubts caused in others concerning the institution of promising. But surely that is a misinterpretation. The wrong is done to the person who should not have suffered from the dislike of others, and to the boy who had a special claim to the promised money. This wrong is present whatever the long-term effects on others.

The rule-utilitarian response misses the real issue. The objection to utilitarian decision-making was that certain special obligations should be included,

and that certain illegitimate preferences should be excluded. These are moral requirements which take precedence over the maximization of utility (whereas the U-agent sees them merely as devices for maximizing utility). But if that was our objection, then it is irrelevant to say, as rule utilitarians do, that obeying promises and discounting prejudices often maximizes long-term utility, or that promises and human rights are even more ingenious devices for maximizing utility than we initially thought. That response confirms, rather than refutes, the criticism that U-agents treat the recognition of special obligations as subject to, rather than prior to, the maximization of utility. Our objection was not that promises are bad devices for maximizing utility, but that they are not such devices at all. This problem cannot be avoided by changing the level at which we apply the principle of utility from acts to rules. The problem, from the point of view of our everyday morality, is in applying the principle of utility itself.

We can make the same point another way. Shifting to rule-utilitarianism may change the *outcome* of the utility calculations, but it does not change the *inputs* into the calculations. The rule utilitarian is still committed to including all preferences, no matter how morally illegitimate they may appear. Focusing on rules rather than acts may make it less likely for illegitimate preferences to win the day, but they still count on a par with all other preferences. Moreover, this has the perverse consequence that the more people enjoy harming others or violating their rights, the less evil is their action. For example, while rule-utilitarianism is unlikely to endorse a lifestyle involving raping and pillaging, it does imply that the enjoyment people take in raping and pillaging counts in the calculus, and the more enjoyment they get, the less the overall wrongness of their action. As Geoffrey Scarre puts it, their enjoyment

seems to *offset some of their evil*: it is a positive quantity in the balance sheet which compensates for some of the suffering of the victims. But it conflicts radically with our ordinary moral convictions to assert that the greater the pleasure a murderous maniac derives from abusing his victim, the smaller the net amount of evil produced by his actions . . . To enjoy the killing makes the killing worse, not better. (Scarre 1996: 155)

Similarly, sadists can offset some of their evil by *sharing* the pleasure involved with other sadists. Rule-utilitarianism is unlikely to condone torturing a child, but it does imply that the torturing of a child is less evil if the torturer shares his pleasure with other sadists—perhaps by inviting an audience, or broadcasting it on the Internet. Such actions may be wrong on rule-utilitarian grounds, but less wrong than if the torturer is the only person who gains pleasure from it.

Or consider the games held in the ancient Roman Colosseum, in which a prisoner of war was torn to shreds by wild animals in front of 50,000 wildly cheering spectators. A clever rule utilitarian can no doubt find reasons why it

would have maximized utility in the long run to give precedence to the rights of a handful of prisoners of war over the blood-lust of the 50,000 spectators.<sup>12</sup> One can reasonably ask whether these clever arguments will still work if we increase the size of the Colosseum to include more people, or if we imagine broadcasting the games on satellite TV so that millions of people around the world can enjoy the spectacle. But again, the real issue here is not the ultimate conclusion utilitarians reach, but the process by which they reach it. On the rule-utilitarian view, the larger the audience for the games, and the more each spectator enjoys it, the less evil it is. On our everyday moral view, by contrast, torturing others becomes more evil, not less, the more people enjoy it.

Some utilitarians would agree with what I have said so far. It is right and proper, they say, to view our attachments and our rights as taking precedence over the pursuit of overall utility. We should accept the everyday view that the harm done to the particular individuals who are cheated or discriminated against is sufficient grounds for demanding that people keep promises and respect rights. We should not be U-agents who decide how to act by making utilitarian calculations, and who view promises as devices for maximizing utility. Instead we should view promises, and other people's rights, as of such towering importance that they are basically invulnerable to the calculus of social interests. In short, we should be non-utilitarians in our moral reasoning. But, they argue, this does not mean that utilitarianism is wrong. On the contrary, the reason why we should be non-utilitarians in our decision-making is precisely that we are more likely to maximize utility that way. A society of non-utilitarians who believe in the intrinsic importance of promises and rights will do better, in terms of maximizing utility, than a society of act or rule utilitarians who view promises and rights as devices for maximizing utility.

This may sound paradoxical. But it raises a true and important point. Utilitarianism is essentially a 'standard of rightness', not a 'decision-procedure' (Brink 1986: 421–7; Railton 1984: 140–6). What defines utilitarianism is the claim that the right act is the one that maximizes utility, not the claim that we should deliberately seek to maximize utility. It is an open question whether we should employ a utilitarian decision-procedure in assessing acts or rules—indeed, this question is itself to be answered by examining the consequences on overall utility of different decision-procedures. And it is quite possible that we would do better in terms of the utilitarian standard of rightness by employing a non-utilitarian decision-procedure. This certainly seems true in regards to our personal attachments—everyone's life is less valuable if we are unable to make commitments in the sort of wholehearted and unconditional way precluded by direct utilitarianism. Hence, it is argued, we should be 'indirect utilitarians' who do not in fact apply utilitarian decision-procedures in our everyday decisions about either acts or rules.

While the distinction between standards of rightness and decision-procedures is sound, if we put too much weight on it, it is not clear why utilitarianism as a standard of rightness should not disappear entirely from our conscious beliefs. Taken to its extreme, indirect utilitarianism could be 'self-defeating'—it might argue for its own elimination from people's thoughts and beliefs (Williams 1973: 135). The world most likely to maximize utility may be one in which no one believes in utilitarianism. A less extreme form of indirect utilitarianism is what Williams calls 'Government House' utilitarianism (Williams and Sen 1982: 16, Williams 1973: 138–40). On this view, a small elite would know that utilitarianism was the right moral theory, and they would employ utilitarian decision-procedures to design utility-maximizing rules or institutions. The vast bulk of the population, however, would not be taught to believe in utilitarianism. They would be taught to view social rules and conventions as intrinsically justified. (This is called 'Government House' utilitarianism since it seems to have been the view of some British colonial officials in India and other British colonies: the British officials would understand that rights are simply ingenious devices for maximizing utility; the natives would be taught to think of rights as intrinsically justified and inviolable.)

This idea of Government House utilitarianism has been widely criticized as elitist, and as violating the democratic norm of 'publicity', according to which the state should be able to publicly justify its actions to its citizens.<sup>13</sup> Most indirect utilitarians, therefore, prefer a model in which everyone shares the same two-level moral outlook. Most of the time, we use non-utilitarian decision-procedures, and view rights and justice as invulnerable to the calculus of utility maximization, but every once in a while (perhaps only in moments of crisis), we all engage in a collective and democratic process of utilitarian decision-making to revise our everyday rules and institutions.

One can question whether this is really a psychologically plausible picture.<sup>14</sup> In any event, it does not yet answer the objections raised above. Consider our everyday view that certain kinds of preferences are unfair, and so should not be given any weight in our moral decision-procedures. It is possible that the utilitarian standard of rightness can justify our adopting such a non-utilitarian decision-procedure. If so, then both sides agree that certain preferences should not be counted. But on our everyday view, the reason why unfair preferences should not be given any weight in our decision-procedure is that they are morally illegitimate—they do not deserve to be counted. For the indirect utilitarian, on the other hand, the reason we do not count unfair preferences is simply that it is counter-productive to do so. Unfair preferences (if rational and informed) are as legitimate as any other preference according to the utilitarian standard of rightness, but we do better in terms of that standard by treating them as illegitimate in our decision-making.

So we have two conflicting explanations for treating certain preferences as illegitimate. To defend utilitarianism, therefore, it is not enough to show that the utilitarian standard of rightness can justify using non-utilitarian decision-procedures. One also must show that this is the right justification. The utilitarian says that the reason why we use non-utilitarian procedures is that they happen to maximize utility. But isn't it more plausible to say that the reason why we use non-utilitarian procedures is simply that we accept a non-utilitarian standard of rightness? Why think there has to be some indirect utilitarian explanation for our non-utilitarian commitments?

Some utilitarians seem to think that if a utilitarian explanation is available for our moral convictions then there is no need to consider any non-utilitarian explanations. But this begs the question. We need some argument for endorsing the utilitarian standard of rightness over alternative standards. Is there any such argument in utilitarian writings? There are in fact two distinct arguments, but I will argue that neither works on its own, and that the plausibility of utilitarianism depends on conflating the two. Once we have examined these arguments, we will see that the problems discussed above stem directly from the utilitarian standard of rightness, and are not substantially affected by how directly or indirectly that standard is applied.

#### 4. TWO ARGUMENTS FOR UTILITY MAXIMIZATION

In this section, I will consider the two main arguments for viewing utility maximization as the standard of moral rightness. As we will see, they generate two entirely different interpretations of what utilitarianism is.

##### (a) Equal consideration of interests

On one interpretation, utilitarianism is a standard for aggregating individual interests and desires. Individuals have distinct and potentially conflicting preferences, and we need a standard that specifies which trade-offs amongst those preferences are morally acceptable, which trade-offs are fair to the people whose welfare is at stake. That is the question which this first interpretation of utilitarianism attempts to answer. One popular answer, found in many different theories, is that each person's interests should be given equal consideration. Each person's life matters equally, from the moral point of view, and hence their interests should be given equal consideration.

Utilitarianism, on this first view of it, accepts this general egalitarian principle. However, the idea of treating people with equal consideration is imprecise, and it needs to be spelled out in more detail if it is to provide a

determinate standard of rightness. One obvious, and perhaps initially appealing, way to spell out that idea is to give equal weight to each person's preferences, regardless of the content of the preferences or the material situation of the person. As Bentham put it, we count everyone for one, no one for more than one. On the first account of utilitarianism, then, the reason that we should give equal weight to each person's preferences is that that treats people as equals, with equal concern and respect.

If we accept this as our standard of rightness, then we will conclude that morally right actions are those that maximize utility. But it is important to note that maximization is not the direct goal of the standard. Maximization arises as a by-product of a standard that is intended to aggregate people's preferences fairly. The requirement that we maximize utility is entirely derived from the prior requirement to treat people with equal consideration. So the first argument for utilitarianism is this:

1. people matter, and matter equally; therefore
2. each person's interests should be given equal weight; therefore
3. morally right acts will maximize utility.

This equal consideration argument is implicit in Mill's claim that 'In the golden rule of Jesus of Nazareth, we read the complete spirit of the ethics of utility. To do as you would be done by, and to love your neighbour as yourself, constitute the ideal perfection of utilitarian morality' (Mill 1968: 16). The argument is more explicitly affirmed by contemporary utilitarians like Harsanyi, Griffin, Singer, and Hare (Harsanyi 1976: 13–14, 19–20, 45–6, 65–7; Griffin 1986: 208–15, 295–301; Hare 1984: 106–12; Singer 1979: 12–23; Haslett 1987: 40–3, 220–2). Hare, in fact, finds it difficult to imagine any other way of showing equal consideration for each person (Hare 1984: 107; cf. Harsanyi: 1976: 35).

##### (b) Teleological utilitarianism

There is, however, another interpretation of utilitarianism. Here maximizing the good is primary, not derivative, and we count individuals equally only because that is the way to maximize value. Our primary duty is not to treat people as equals, but to bring about valuable states of affairs. People, as Williams puts it, are just viewed as *locations* of utilities, or as causal levers for the 'utility network'. The 'basic bearer of value for Utilitarianism is the *state of affairs*' (Williams 1981: 4). Utilitarianism, on this view, is primarily concerned not with persons, but with states of affairs. Rawls calls this a 'teleological' theory, which means that the right act is defined in terms of maximizing the good, rather than in terms of equal consideration for individuals (Rawls 1971: 24).

This second interpretation is a genuinely distinct form of utilitarianism, not simply a different way of describing the same theory. Its distinctiveness

becomes clear if we look at utilitarian discussions of population policy. Derek Parfit asks whether we morally ought to double the world's population, even if it means reducing each person's welfare by almost half (since that will still increase overall utility). He thinks that a policy of doubling the population is a genuine, if somewhat repugnant, conclusion of utilitarianism.

Indeed, we should not stop with simply doubling the population. A world with 100 billion people, each of whom has a life barely worth leading, might well contain more overall utility than a world of 5 billion people, each of whom has a very high quality of life. Compare two possible worlds: world A, our world, containing 5 billion people each of whom has an average utility of 18 units, and world B, containing 100 billion, each of whose well-being has been reduced to one unit (see Fig. 1).

In World B, each person's life has become miserable—barely better than being dead—yet the overall amount of utility has increased from 90 to 100 billion units. Utilitarians, according to Parfit, should seek to maximize the total amount of utility in the world, no matter what its impact on the utility of existing individuals, and hence prefer World B (Parfit 1984: 388).

But this need not be the conclusion if we view utilitarianism as a theory of treating people as equals. Non-existent people do not have claims—we do not have a moral duty *to them* to bring them into the world. As John Broome

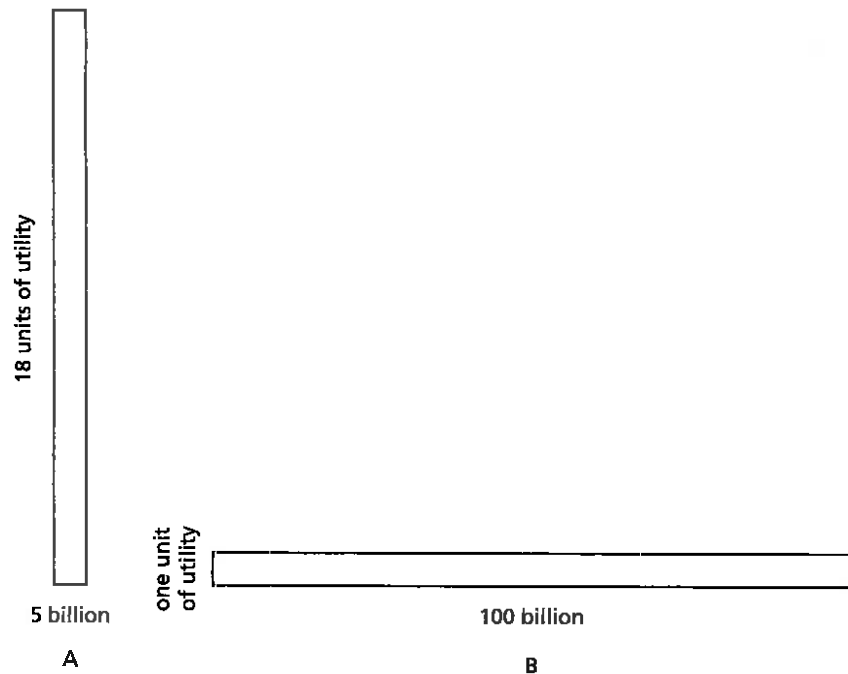


Figure 1 Parfit's repugnant conclusion

notes, 'one cannot owe anyone a duty to bring her into existence, because failing in such a duty would not be failing anyone' (Broome 1991: 92). So what is the duty here, on the second interpretation? The duty is to maximize value, to bring about valuable states of affairs, even if the effect is to make all existing persons worse off than they otherwise would have been.

The distinctness of this second interpretation is also apparent in Thomas Nagel's discussion. He demands that we add a 'deontological' constraint of equal treatment onto utilitarianism, which he thinks is concerned with selecting the 'impersonally best outcome' (Nagel 1986: 176). Nagel says we must qualify our obligation to maximize the good with the obligation to treat people as equals. Obviously his demand only makes sense with reference to the second interpretation of utilitarianism, according to which the fundamental duty is not to aggregate individual preferences fairly, but to bring about the most value in the world. For on the first interpretation, utilitarianism is already a principle of moral equality; if it fails as a principle of equal consideration, then the whole theory fails, for there is no independent commitment to the idea of maximizing utility.

This second interpretation stands the first interpretation on its head. The first defines the right in terms of treating people as equals, which leads to the utilitarian aggregation standard, which happens to maximize the good. The second defines the right in terms of maximizing the good, which leads to the utilitarian aggregation standard, which as a mere consequence treats people's interests equally. As we have seen, this inversion has important theoretical and practical consequences.

So we have two independent, and indeed conflicting, paths to the claim that utility ought to be maximized. Which is the fundamental argument for utilitarianism? Up to this point, I have implicitly relied on the first view—that is, utilitarianism is best viewed as a theory of how to respect the moral claim of each individual to be treated as an equal. Rawls, however, says that utilitarianism is fundamentally a theory of the second sort—i.e. one which defines the right in terms of maximizing the good (Rawls 1971: 27). But there is something bizarre about that second interpretation. For it is entirely unclear why maximizing utility, as our direct goal, should be considered a *moral* duty. To whom is it a duty? Morality, in our everyday view, is a matter of interpersonal obligations—the obligations we owe to each other. But to whom do we owe the duty of maximizing utility? It cannot be to the maximally valuable state of affairs itself, for states of affairs do not have moral claims. Perhaps we have a duty to those people who would benefit from the maximization of utility. But if that duty is, as seems most plausible, the duty to treat people with equal consideration, then we are back to the first interpretation of utilitarianism as a way of treating people as equals. Maximizing utility is now just a by-product, not the ultimate ground of the theory. And then we need not double the

population, since we have no obligation to conceive those who would constitute the increased population.<sup>15</sup>

If we nonetheless accept that maximizing utility is itself the goal, then it is best seen as a non-moral ideal, akin in some ways to an aesthetic ideal. The appropriateness of this characterization can be seen by looking at the other example Rawls gives of a teleologist, namely Nietzsche (Rawls 1971: 25). The good which Nietzsche's theory seeks to maximize (e.g. creativity) is available only to the special few. Others are useful only in so far as they promote the good of the special few. In utilitarianism, the value being maximized is more mundane, something that every individual is capable of partaking in or contributing to (although the maximizing policy may result in the sacrifice of some people). This means that in utilitarian teleology, unlike Nietzsche's, every person's preferences must be given some weight. But in neither case is the fundamental principle to treat people as equals. Rather it is to maximize the good. And in both cases, it is difficult to see how this can be viewed as a moral principle. The goal is not to respect *people*, for whom certain things are needed or wanted, but rather to respect the *good*, to which certain people may or may not be useful contributors. If people have become the means for the maximization of the good, morality has dropped out of the picture, and a non-moral ideal is at work. A Nietzschean society may be aesthetically better, more beautiful, but it is not morally better (Nietzsche himself would not have rejected this description—his theory was 'beyond good and evil'). If utilitarianism is interpreted in this teleological way then it too has ceased to be a moral theory.

I said earlier that one of utilitarianism's attractions was its secular nature—for utilitarians, morality matters because human beings matter. But that attractive idea is absent from this second interpretation, whose moral point is quite obscure. Humans are viewed as potential producers or consumers of a good, and our duties are to that good, not to other people. That violates our core intuition that morality matters because humans matter. In fact, few people have endorsed utilitarianism as a purely teleological theory, without appealing at all to the ideal of equal respect for persons (G. E. Moore's *Ethics* is one prominent exception). Utilitarianism simply ceases to have any attraction if it is cut off from that core intuition.

If utilitarianism is best seen as an egalitarian doctrine, then there is no independent commitment to the idea of maximizing welfare. The utilitarian has to admit that we should use the maximizing standard only if that is the best account of treating people as equals. This is important, because much of the attraction of utilitarianism depends on a tacit mixing of the two justifications.<sup>16</sup> Utilitarianism's intuitive unfairness would quickly disqualify it as an adequate account of equal consideration, were it not that many people take its maximizing feature as an additional, independent reason to endorse it.

Utilitarians tacitly appeal to the good-maximization standard to deflect intuitive objections to their account of equal consideration. Indeed, it may seem to be a unique strength of utilitarianism that it can mix these two justifications. Unfortunately, it is incoherent to employ both standards in the same theory. One cannot say that morality is fundamentally about maximizing the good, while also saying that it is fundamentally about respecting the claim of individuals to equal consideration. If utilitarians were held to one or other of the standards, then their theory would lose much of its attractiveness. Viewed as a maximizing-teleological theory, it ceases to meet our core intuitions about the point of morality; viewed as an egalitarian theory, it leads to a number of results which conflict with our sense of what it is to treat people as equals, as I now hope to show in a more systematic way.

## 5. INADEQUATE CONCEPTION OF EQUALITY

If we are to treat utilitarianism as a plausible political morality, then we must interpret it as a theory of equal consideration. That may seem strange, given the inegalitarian acts utilitarianism might justify—e.g. depriving disliked people of their liberty. But we need to distinguish different levels at which equality can be a value. While utilitarianism may have unequal effects on people, it can nonetheless claim to be motivated by a concern for treating people as equals. Indeed, Hare asks, if we believe that people's essential interest is the satisfaction of their informed preferences, and that everyone is to be given equal consideration, then what else can we do except give equal weight to each person's preferences, everyone counting for one, no one for more than one (Hare 1984: 106)?

But while utilitarianism seeks to treat people as equals, it violates many of our intuitions about what it genuinely means to treat people with equal consideration. It is possible that our anti-utilitarian intuitions are unreliable. I will argue, however, that utilitarianism has misinterpreted the ideal of equal consideration for each person's interests, and, as a result, it allows some people to be treated as less than equals, as means to other people's ends.

Why is utilitarianism inadequate as an account of equal consideration? Utilitarians assume that every source of happiness, or every kind of preference, should be given the same weight, if it yields equal utility. I will argue that an adequate account of equal consideration must distinguish different kinds of preferences, only some of which have legitimate moral weight.

### (a) External preferences

One important distinction amongst kinds of preferences is that between 'personal' and 'external' preferences (Dworkin 1977: 234). Personal preferences are

preferences about the goods, resources, and opportunities etc. one wants available to oneself. External preferences concern the goods, resources, and opportunities one wants available to others. External preferences are sometimes prejudiced. Someone may want blacks to have fewer resources because he thinks them less worthy of respect. Should this sort of external preference be counted in the utilitarian calculus? Does the existence of such preferences count as a moral reason for denying blacks those resources?

As we have seen, indirect utilitarians argue that there are circumstances where we would be better off, in utilitarian terms, by excluding such preferences from our everyday decision-procedures. But the question I want to consider here is whether these preferences should be excluded more systematically, by excluding them from our standard of rightness. And I want to consider whether utilitarianism's own deepest principle provides grounds for not according external preferences any moral weight in its standard of rightness. The deepest principle, as we have seen, is an egalitarian one. Each person has an equal moral standing, each person matters as much as any other—that is why each person's preferences should count in the calculus. But if that is why we are attracted to utilitarianism, then it seems inconsistent to count external preferences. For if external preferences are counted, then what I am rightfully owed depends on how others think of me. If they think I am unworthy of equal concern, then I will do less well in the utilitarian aggregation. But utilitarians cannot accept that result, because utilitarianism is premised on the view that everyone ought to be treated as equals.

If we believe that everyone is to be treated as equals, then it offends our deepest principles to allow some people to suffer because others do not want them treated as equals. As Dworkin puts it, *inegalitarian external preferences 'are on the same level—purport to occupy the same space—as the utilitarian theory'*. Hence utilitarianism 'cannot accept at once a duty to defeat the false theory that some people's preferences should count for more than other people's and a duty to strive to fulfill the [external] preferences of those who passionately accept that false theory, as energetically as it strives for any other preferences' (Dworkin 1985: 363). The very principle that tells us to count equally every person's preferences in our standard of rightness also tells us to exclude those preferences which deny that people's preferences are to count equally. To paraphrase Harsanyi, utilitarians should be 'conscientious objectors' when faced with such preferences (Harsanyi 1977a: 62; Goodin 1982: 93–4).

### (b) Selfish preferences

A second kind of illegitimate preference involves the desire for more than one's own fair share of resources. I will call these 'selfish preferences', since they ignore the fact that other people need the resources, and have legitimate

claims to them. As with *inegalitarian external preferences*, selfish preferences are often irrational and uninformed. But satisfying selfish preferences will sometimes generate genuine utility. Should such preferences, if rational, be included in the utilitarian standard of rightness?

Utilitarians will object to the way I have phrased the question. As we have seen, utilitarians deny that there is such a thing as a fair share (and hence a selfish preference) independently of utilitarian calculations. For utilitarians, a fair distribution just is one that maximizes utility, and so no preference can be identified as selfish prior to utility calculations. So it begs the question against utilitarianism to assume that we can identify such things as selfish preferences prior to utilitarian calculations. But we can ask whether the utilitarian's own deepest principle provides grounds for adopting a theory of fair shares that enables us to identify and exclude selfish preferences from our standard of rightness.

This issue is discussed in a recent debate between Hare and John Mackie. Hare, like most utilitarians, believes that all rational preferences should be included in utility aggregation, even those that seem unfair. Even if I have a massive amount of resources, while my neighbour has very little, if I covet my neighbour's resources, then my desire must be included in the calculation. And if the calculations work out in my favour, perhaps because I have many friends who would share in my enjoyment, then I should get those resources. No matter how much I already have, my desire for more resources continues to count equally, even when the resources I want must come from someone with very little.

Why should utilitarians count such preferences? Hare believes that the principle of equal consideration requires it. According to Hare, the best way to interpret that egalitarian principle is to use the following mental test: we put ourselves in other people's shoes, and try to imagine how our actions affect them. And we should do this for everyone affected by our actions. We take the viewpoint of each person and treat it as being equally important as our own viewpoint, equally worthy of concern. Indeed, Hare says, we should treat these other viewpoints *as* our own viewpoint. This ensures that we are showing equal consideration for each person. If we have, in this way, put ourselves in everyone else's shoes, then we should choose that action which is best for 'me', where 'me' here means all of the 'me's', i.e. all of the different viewpoints I am now considering as equally my own. If I try to choose what is best for all my different selves, I will choose that action which maximizes the preference satisfaction of all these 'selves'. So, Hare claims, the utilitarian aggregation criterion follows naturally from this intuitive model of equal consideration. If I treat each person's interests as mattering equally, by imagining that their viewpoint is in fact one of my own, then I will adopt utilitarian principles (Hare 1984: 109–10; cf. 1982: 25–7).



Hare thinks that this is the only rational way of showing equal concern for people. But as Mackie notes, there are other possibilities, even if we accept Hare's claim that we treat people as equals by putting ourselves in their shoes, and treating each of these different selves as equally important. Rather than maximize preference satisfaction amongst all these selves, we might show our concern for them by guaranteeing each 'a fair go' in life, i.e. guarantee each an adequate level of resources and liberties. Or we might, when successively occupying these different positions, do what is best for the least well off, or provide each an equal share of the available resources and liberties. These are all different conceptions of what the abstract notion of equal consideration requires (Mackie 1984: 92).

How can we decide between these different ways of showing equal consideration? Utilitarians point out that their view may also lead to an egalitarian distribution of resources. People who lack resources will, in general, get more utility out of each additional resource than those who already have many resources. Someone who is starving is sure to get more utility from a piece of food than someone who is already well supplied with food (Hare 1978: 124–6; Brandt 1959: 415–20; Goodin 1995: 23). We can represent this graphically (see Fig. 2). If we take \$10 from a rich person at point R (moving them down to point R<sub>1</sub>), and give it to a poor person at point P (moving them up to point P<sub>1</sub>), we will increase overall utility—P gains much more in utility than R loses.

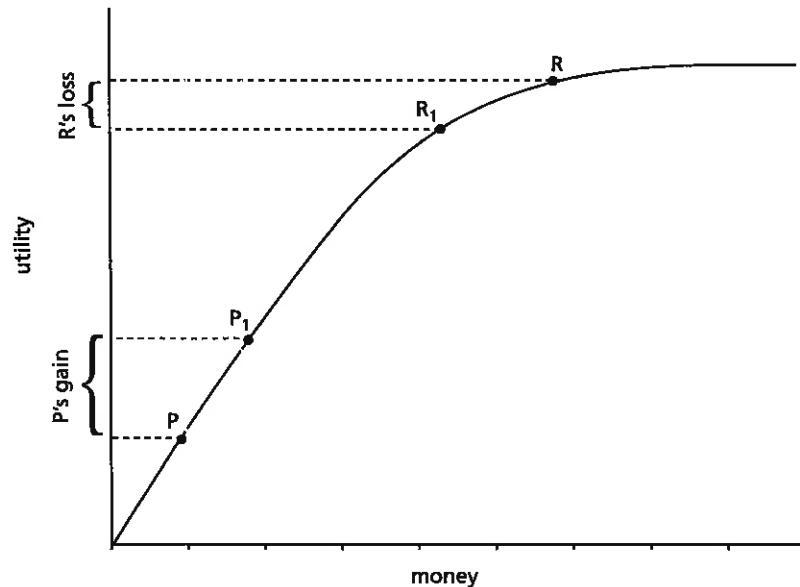


Figure 2 Declining marginal utility

So both sides can agree to start with a roughly equal distribution of resources. However, Hare and Mackie conceive this initially equal distribution in very different ways. For Mackie, so long as everyone else has their fair share of resources, then the resources initially allotted to me are mine—i.e. no one else has any legitimate claim of justice over them. Some people who already have their fair share may also want some of my share. But that is not important, morally speaking. Their preferences have no moral weight. They are selfish preferences, since they fail to respect my claim to a fair share. On Mackie's view, the state should secure each person's share of resources, and not allow them to be taken away just because other people have selfish preferences for what is rightfully someone else's. The best conception of equal consideration would exclude such selfish preferences.

For Hare, on the other hand, the resources initially distributed to me are not really mine in the same way. They are mine unless or until someone else can make better use of them, where 'better' means more productive of overall utility. Hare thinks this proviso for taking away my share is required by the same value that led the government initially to give it to me, i.e. an equal concern for each person's goals. If we care equally about people's goals, then it is right to redistribute resources whenever we can satisfy more goals by so doing.

Do we have any reason to choose one of these conceptions of equal consideration over the other? We need to look more closely at the kinds of preferences that would be involved in Hare's redistribution. Let us assume that I have my fair share, as does everyone else, and that we are in an affluent society, so that this share includes a house and lawn. Everyone else on my block plants a flower garden, but they would like my lawn left open as a public space for children to play on, or to walk dogs on. I, however, want my own garden. The desires of others to use my lawn as a public space may well outweigh, in terms of overall utility, my desire to have a garden. Hare thinks it is right, therefore, to sacrifice my desire for the greater desires of others.

If it is morally wrong for me to insist on having a garden, we need to know *who is wronged*. If my sacrifice is required to treat people as equals, who is treated as less than an equal if I disallow the sacrifice? Hare's answer is that the other members of the block are not treated as equals if their preferences are not allowed to outweigh my desire. But surely that is implausible, since they already have their own yard, their own fair share of resources. According to Hare, my neighbours' desire to decide how to use my resources, as well as their own, is a legitimate preference which grounds a moral claim. But isn't it more accurate to describe such a preference as simply selfish? Why should my neighbours suppose that the idea of equal concern gives them any claim over my share of resources? If they already have their own lawn, then I am not treating them unjustly in saying that my preference concerning my lawn

outweighs or pre-empts their preferences. I still respect them as equals since I make no claim on the resources they have to lead their lives. But they do not respect me as an equal when they expect or demand that I give up my share of resources to satisfy their selfish desire to have more than their fair share.

This points to an important component of our everyday sense of what it means to treat people as equals—namely, we should not expect others to subsidize our projects at the expense of their own. Perhaps my friends and I have expensive tastes—we like to eat caviar and play tennis all day. To expect others to give up their fair share of resources to support our taste, no matter how happy it makes us, is selfish. If I already have my share of resources, then to suppose that I have a legitimate moral claim to someone else's resources, just because it will make me happier, is a failure to show equal concern for others. If we believe that others should be treated as equals, then we will exclude such selfish preferences from the utilitarian calculus.

So the very principle which supported an initially equal distribution of resources also argues for securing that distribution. Hare's proviso—that the initial distribution be subject to utility-maximizing redistribution—undermines, rather than extends, the point of the initial distribution. Hare's idea of treating other people's interests as my own when engaged in moral reasoning is not necessarily a bad one. It is one way of rendering vivid the idea of moral equality (we will look at other such devices in the next chapter). But the equal concern he seeks to promote is not achieved by treating other people's preferences as constituting equal claims on all of our actions and resources. Rather, equality teaches us how much by way of resources we have to pursue our projects, and how much is rightfully left for others. Equal concern is shown by ensuring that others can claim their own fair share, not by ensuring that they have equal weight in determining the use of my share. Securing people's fair shares, rather than leaving them subject to selfish preferences, is the better spelling out of the equal concern that Hare seeks.

This, according to Rawls, is a fundamental difference between his account of justice and the utilitarians'. For Rawls, it is a defining feature of our sense of justice that 'interests requiring the violation of justice have no value', and so the presence of illegitimate preferences 'cannot distort our claims upon one another' (Rawls 1971: 31, 450, 564). Justice 'limits the admissible conceptions of the good, so that those conceptions the pursuit of which violate the principles of justice are ruled out absolutely: the claims to pursue inadmissible conceptions have no weight at all'. Because unfair preferences 'never, so to speak, enter into the social calculus', people's claims 'are made secure from the unreasonable demands of others'. For utilitarians, on the other hand, 'no restrictions founded on right and justice are imposed on the ends through which satisfaction is to be achieved' (Rawls 1982*b*: 184, 171*n.*, 170, 182).

We can now see why utilitarianism fails to recognize special relationships,

or to exclude illegitimate preferences. In each case, utilitarianism is interpreting equal consideration in terms of the aggregation of pre-existing preferences, whatever they are for, even if they invade the rights or commitments of others. But our intuitions tell us that equality should enter into the very formation of our preferences. Part of what it means to show equal consideration for others is taking into account what rightfully belongs to them in deciding on one's own goals in life.<sup>17</sup> Hence prejudiced and selfish preferences are excluded from the start, for they already reflect a failure to show equal consideration. However, if my goals do respect other people's rightful claims, then I am free to pursue special relationships, even if some other act maximizes utility. If my plans respect the teachings of equality, then there is nothing wrong with giving priority to my family or career. This means that my day-to-day activities will show unequal concern—I will care more about helping my friends, or the causes I am committed to, than about helping the goals of other people. That is part of what it means to have friends and causes. And that is entirely acceptable, so long as I respect the claims of others concerning the pursuit of their projects.

If we think about the values that motivate utilitarianism, the values which give it its initial plausibility, we will see that it must be modified. Utilitarianism is initially attractive because human beings matter and matter equally. But the goal of equal consideration that utilitarians seek to implement is best implemented by an approach that includes a theory of fair shares. Such a theory would exclude prejudiced or selfish preferences that ignore the rightful claims of others, but would allow for the kinds of special commitments that are part of our very idea of leading a life. These modifications do not conflict with the general principle of consequentialism, but rather stem from it. They are refinements of the general idea that morality should be about the welfare of human beings. Utilitarianism has simply oversimplified the way in which we intuitively believe that the welfare of others is worthy of moral concern.

In defending the importance of rights which protect people from utilitarian aggregation, Rawls and Mackie do not dispute the moral importance of consequences. As Rawls notes, 'all ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy' (Rawls 1971: 30). Rawls, Mackie, and other 'rights-based' theorists simply build concern for consequences into their theories at a different and indeed earlier stage than utilitarianism. They argue that morality requires us to take the consequences for others into account in the very formation of our preferences, not just in the aggregation of those preferences.

As we have seen, indirect utilitarians claim that our intuitive commitment to non-utilitarian decision-procedures does not undermine utilitarianism as a standard of rightness, since we can give a utilitarian justification for adopting non-utilitarian procedures. But that response will not work here, for my

argument concerns utilitarianism as a standard of rightness. My claim is that the very reason utilitarians give for basing their standard of rightness on the satisfaction of people's preferences is also a reason to exclude external and selfish preferences from that standard. This is an objection to the theory's principles, not to the way those principles get applied in decision-procedures.

Commentators who endorse these sorts of modifications of utilitarianism often describe the resulting theory as a balance between the values of utility and equality, or a compromise between consequentialism and deontology (e.g. Raphael 1981: 47–56; Brandt 1959: ch. 16; Hospers 1961: 426; Rescher 1966: 59). That is not what I have argued. Rather, the modifications are needed to provide a better spelling out of the ideal of equal consideration which utilitarianism itself appeals to.

It is worth pausing to consider the kind of argument that I have just presented, since it expresses, I believe, one basic form of political argument. As I mentioned in the Introduction, the idea of equality is often said to be the basis of political morality. Both Hare's utilitarianism and Mackie's 'right to a fair go' appeal to the idea that each person is entitled to equal consideration. But they do not give an equally compelling account of that idea. Our intuitions tell us that utilitarianism fails to ensure that people are treated as equals, since it lacks a theory of fair shares.

This might suggest that political theorizing is a matter of correctly deducing specific principles from this shared premiss of moral equality. Political argument, then, would primarily be a matter of identifying mistaken deductions. But political philosophy is not like logic, where the conclusion is meant to be already fully present in the premisses. The idea of moral equality is too abstract for us to be able to deduce anything very specific from it. There are many different and conflicting kinds of equal treatment. Equality of opportunity, for example, may produce unequal income (since some people have greater talents), and equal income may produce unequal welfare (since some people have greater needs). All of these particular forms of equal treatment are logically compatible with the idea of moral equality. The question is which form of equal treatment best captures that deeper ideal of treating people as equals. That is not a question of logic. It is a moral question, whose answer depends on complex issues about the nature of human beings and their interests and relationships. In deciding which particular form of equal treatment best captures the idea of treating people as equals, we do not want a logician, who is versed in the art of logical deductions. We want someone who has an understanding of what it is about humans that deserves respect and concern, and of what kinds of activities best manifest that respect and concern.

The idea of moral equality, while fundamental, is too abstract to serve as a premiss from which we deduce a theory of justice. What we have in political

argument is not a single premiss and then competing deductions, but rather a single concept and then competing conceptions or interpretations of it. Each theory of justice is not *deduced from* the ideal of equality, but rather *aspires to* it, and each theory can be judged by how well it succeeds in that aspiration. As Dworkin puts it, when we instruct public officials to act in accordance with the concept of equality, we 'charge those whom [we] instruct with the responsibility of developing and applying their own conception . . . That is not the same thing, of course, as granting them a discretion to act as they like; it sets a standard which they must try—and may fail—to meet, because it assumes that one conception is superior to another' (Dworkin 1977: 135).<sup>18</sup> However confident we are in a particular conception of equality, it must be tested against competing conceptions to see which best expresses or captures the concept of equality.

This is the kind of argument I have tried to give against utilitarianism. We can see the weakness in utilitarianism as a conception of equality by comparing it to a conception which guarantees certain rights and fair shares of resources. When we compare these two conceptions, utilitarianism seems implausible as an account of moral equality, at odds with our intuitions about that basic concept. But its implausibility is not a matter of logical error, and the strength of a theory of fair shares isn't a matter of logical proof. This may be unsatisfying to those accustomed to more rigorous forms of argument. But if the egalitarian suggestion is correct—if each of these theories is aspiring to live up to the ideal of treating people as equals—then this is the form that political argument must take. To demand that it achieve logical proof simply misunderstands the nature of the exercise. Any attempt to spell out and defend our beliefs about the principles which should govern the political community will take this form of comparing different conceptions of the concept of equality.

## 6. THE POLITICS OF UTILITARIANISM

What are the practical implications of utilitarianism as a political morality? I have noted the danger that utilitarianism could justify sacrificing the weak and unpopular members of the community for the benefit of the majority. But utilitarianism has also been used to attack those who hold unjust privileges at the expense of the majority. Indeed, utilitarianism, as a self-conscious political and philosophical movement, arose as a radical critique of English society. The original utilitarians were 'Philosophical Radicals' who believed in a complete rethinking of English society, a society whose practices they believed were the product not of reason, but of feudal superstition. Utilitarianism, at that time, was identified with a progressive and reform-minded

political programme—the extension of democracy, penal reform, welfare provisions, etc.

Contemporary utilitarians, on the other hand, are ‘surprisingly conformist’—in fact they seem keen to show that utilitarianism leaves everything as it is (Williams 1972: 102). As Stuart Hampshire noted, British utilitarianism ‘set out to do good in the world’, and

succeeded in large part over many years in this aim. . . . The utilitarian philosophy, before the First World War and for many years after it . . . was still a bold, innovative, even a subversive doctrine, with a record of successful social criticism behind it. I believe that it is losing this role, and that it is now an obstruction. (Quoted in Goodin 1995: 3)

To be sure, some utilitarians continue to claim that utilitarianism requires a radical critique of the arbitrary and irrational aspects of everyday morality (e.g. Singer 1979). But utilitarianism no longer forms a coherent political movement, and tends if anything to defend the status quo.

What explains this increasing conservatism? I think there are two main reasons. The first is the increasing recognition of the difficulty in actually applying utilitarian principles. Whereas the original utilitarians were willing to judge existing social codes at the altar of human well-being, many contemporary utilitarians argue there are good utilitarian reasons to defer to everyday morality. It may seem that we can increase utility by making exceptions to a rule of everyday morality, but there are utilitarian reasons for sticking to good rules under all circumstances. The gains of new rules are uncertain, whereas existing conventions have proven value (having survived the test of cultural evolution), and people have formed expectations around them. And even if it seems that the everyday rule is not a good one in utilitarian terms, there are utilitarian reasons for not evaluating rules in terms of utility. Acting directly on utilitarian grounds is counter-productive, for it encourages a contingent and detached attitude towards what should be wholehearted personal and political commitments. Moreover, it is difficult to predict the consequences of our actions, or to measure these consequences even when known. Hence our judgements about what maximizes utility are imperfect, and attempts to rationalize social institutions are likely to cause more harm than good.

As a result, modern utilitarians downplay the extent to which utilitarianism should be used as a critical principle, or as a principle of political evaluation at all.<sup>19</sup> Some utilitarians say we should only resort to utilitarian reasoning when our everyday precepts lead to conflicting results; others say that the best world, from a utilitarian point of view, is one in which no one ever reasons in an explicitly utilitarian manner. Williams claims that this sort of utilitarianism is self-defeating—it argues for its own disappearance. This is not self-defeating

in the technical sense, for it does not show that the morally right action is not, after all, the one that maximizes utility. But it does show that utilitarianism is no longer being offered as the correct language for political debate. Politics should be debated in the non-utilitarian language of everyday morality—the language of rights, personal responsibilities, the public interest, distributive justice, etc. Utilitarianism, on some modern views of it, leaves everything as it is—it stands above, rather than competes with, everyday political decision-making.

There is another reason why utilitarianism has become more conservative. Utilitarianism arose in Britain at a time when much of society was still organized to benefit a small, privileged elite at the expense of the (rural and working-class) *majority*. This elitist social structure was often justified in terms of some ideologically biased conception of tradition, nature, or religion. The fundamental political disputes were about whether or not to reform the system to enhance the rights of the majority. In these circumstances, utilitarianism’s commitment to secularism and maximization meant that it sided clearly with the historically oppressed majority against the privileged elite.

In contemporary liberal democracies, however, the fundamental political questions are different. The majority (or at least its male members) has long since acquired its basic civil and political rights. Starting with the civil rights movements in the 1950s and 1960s, many of the burning political questions have centred on the rights of historically oppressed *minorities*—such as African-Americans, gays, indigenous peoples, or people with disabilities. Moreover, these rights are typically asserted against the majority—i.e. they are intended to force the majority to accept policies that are not desired by, or in the interests of, the majority. In these cases, utilitarianism no longer offers such clear or unambiguous direction. The minority in question may be both small—perhaps only 2–5 per cent of the population—and unpopular. Many members of the majority are prejudiced against such minorities, and even if not, the majority has historically supported and benefited from the oppression of various minorities. The majority has enriched itself by dispossessing indigenous peoples, for example. According land rights to indigenous peoples, or accessibility rights to the disabled, may involve significant financial costs to members of the majority, and force them to give up cherished traditions and practices that excluded the minority.

In these circumstances, it is far from clear what utilitarianism recommends. If we simply count up votes or measure public opinion, we may well find that opponents of gay rights outnumber the supporters. Or if we count up who gains or loses from indigenous land rights, we may well find that more people lose than gain from these rights. A simple application of utilitarianism would seem to side with the majority against the minority seeking its rights. Of course, as we have seen, utilitarians have various reasons for saying that in the

long term, everyone benefits when the rights of even small and unpopular minorities are protected against the prejudices or economic interests of the majority. We need to weigh the short-term desire or interest of the majority in oppressing or neglecting a particular minority against the long-term interests of maintaining stable and functioning institutions. But these are complex and speculative questions on which utilitarians themselves disagree.

In short, when the question is whether to defend an oppressed majority against a small privileged elite, utilitarianism gives us a clear and progressive answer. But when the question is whether to defend an oppressed minority against a large privileged majority, utilitarianism gives us vague and conflicting answers, depending on how we identify and weigh short-term and long-term effects. The problem is that 'the winds of utilitarian argumentation blow in too many directions' (Sher 1975: 159). This problem applies in virtually all areas of public policy. For example, while some utilitarians argue that utility is maximized by massive redistribution of wealth, due to the declining marginal utility of money, others defend laissez-faire capitalism because it creates more wealth. This is not just a question of predicting how different economic policies fare in terms of an agreed-upon scale of utility. It is also a question about how to define the scale—what is the relationship between economic goods and other components of the human good (leisure, community, etc.)? It is also a question of the role of utility calculations themselves—how reliably can we determine overall utility, and how important are established conventions? Given these disagreements about how and when to measure utility, utilitarianism is bound to yield fundamentally opposed judgements.

I do not mean to suggest that all these positions are equally plausible (or that these problems are not also found in non-utilitarian theories). The confidence and unanimity that the original utilitarians had in their political judgements was often the result of an oversimplified view of the issues, and a certain amount of indeterminacy is unavoidable in any theory once we recognize the complexity of the empirical and moral issues involved. Modern utilitarians are right to insist that utility is not reducible to pleasure, and that not all kinds of utility are measurable or commensurable, and that it is not always appropriate even to try to measure these utilities. However, the price of this added sophistication is that utilitarianism does not immediately identify any set of policies as distinctly superior. Modern utilitarianism, despite its radical heritage, no longer defines a distinctive political position.

## GUIDE TO FURTHER READING

The most famous statements of utilitarianism remain those of its nineteenth-century founders, particularly Jeremy Bentham, John Stuart Mill, and Henry Sidgwick. Indeed,

much of the literature on utilitarianism even today consists of commentaries on these authors. For these classical statements, see Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns and H. L. A. Hart (Athlone Press, 1970, 1st pub. 1823); J. S. Mill, *Utilitarianism, Liberty, Representative Government*, ed. A. D. Lindsay (J. M. Dent and Sons, 1968, 1st pub. 1863); and Henry Sidgwick, *The Methods of Ethics* (Hackett, 1981, 1st pub. 1874). For contemporary commentaries, see David Lyons (ed.), *Mill's Utilitarianism: Critical Essays* (Rowman and Littlefield, 1997), Roger Crisp (ed.), *Routledge Philosophy Guidebook to Mill on Utilitarianism* (Routledge, 1997); Ross Harrison (ed.), *Bentham* (Routledge, 1999); Bart Schultz (ed.), *Essays on Sidgwick* (Cambridge University Press, 1992).

Much of the literature for and against utilitarianism treats it as a general theory of ethics or personal morality, intended to guide or evaluate our personal conduct and choices. For influential contemporary defences of utilitarian ethics, see James Griffin, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford University Press, 1986); David Lyons, *Forms and Limits of Utilitarianism* (Oxford University Press, 1965); Richard Brandt, *A Theory of the Right and the Good* (Oxford University Press, 1979), and *Morality, Utilitarianism and Rights* (Cambridge University Press, 1992); R. M. Hare, *Moral Thinking* (Oxford University Press, 1981). For an attempt to apply utilitarianism to a wide range of practical problems, from euthanasia to Third World poverty to animal rights, see Peter Singer, *Practical Ethics* (Cambridge University Press, 1993).

Relatively less has been written defending utilitarianism as a specifically political morality for the evaluation of political institutions and public policies. For two important exceptions, see Robert Goodin, *Utilitarianism as a Public Philosophy* (Cambridge University Press, 1995); and James Bailey, *Utilitarianism, Institutions, and Justice* (Oxford University Press, 1997).

Whether offered as a doctrine of personal ethics or political institutions, utilitarianism has been subject to withering critiques. One of the earliest, and still powerful, critiques is by Bernard Williams in J. J. C. Smart and B. Williams (eds.), *Utilitarianism: For and Against* (Cambridge University Press, 1973). Other important critiques (and replies) can be found in Amartya Sen and Bernard Williams (eds.), *Utilitarianism and Beyond* (Cambridge University Press, 1982); Raymond Frey (ed.), *Utility and Rights* (University of Minnesota Press, 1984); and Lincoln Allison (ed.), *The Utilitarian Response: The Contemporary Viability of Utilitarian Political Philosophy* (Sage, 1990).

Two introductory surveys of these debates are Geoffrey Scarre, *Utilitarianism* (Routledge, 1996) in the Routledge 'Problems of Philosophy' series; and William Shaw, *Contemporary Ethics: Taking Account of Utilitarianism* (Blackwell, 1998). Many of the most important readings are excerpted in Jonathan Glover (ed.), *Utilitarianism and its Critics* (Macmillan, 1990).

For those wishing to keep up with new developments in the field, the journal *Utilitas* specializes in the study of utilitarianism, and *Economics and Philosophy* often contains debates between utilitarians and their critics. There are also a couple of helpful websites devoted to utilitarianism. The first is the website of the 'Bentham Project' at University College London, which includes the site for *Utilitas* and the International Society for Utilitarian Studies ([www.ucl.ac.uk/Bentham-Project/](http://www.ucl.ac.uk/Bentham-Project/)). The

second is 'Utilitarian Resources' ([www.utilitarianism.com](http://www.utilitarianism.com)). Both include extensive bibliographies and on-line texts.

## NOTES

1. This common slogan is misleading because it contains two distinct maximands—'greatest happiness' and 'greatest number'. It is impossible for any theory to contain a double maximand, and any attempt to implement it quickly leads to an impasse (e.g. if the two possible distributions are 10:10:10 and 20:20:0, then we cannot produce both the greatest happiness and the happiness of the greatest number). See Griffin 1986: 151–4; Rescher 1966: 25–8.

2. For discussions of adaptive preferences, see Elster 1982*b*; 1983*b*; Barry 1989*b*; Sunstein 1991; Sunstein 1997: chs. 1–2. For applications to gender issues, see Sunstein 1999; Okin 1999; Nussbaum 2000. This is related, of course, to the Marxist theory of false consciousness, according to which workers have been socialized in such a way as to be unable to see their real interest in socialism.

3. Of course, while I might prefer A if informed, it does not follow that A provides me with any benefit in my current uninformed state. This complicates the informed preference account of utility, but does not subvert it. What promotes my well-being is distinct from satisfying my existing preferences, even if it is also distinct from satisfying my ideally informed preferences (Griffin 1986: 11–12; 32–3). It is possible, however, that a full development of this account would bring it close to what is sometimes called an 'Objective List' theory (Parfit 1984: 493–502).

4. For discussions of 'incommensurability', and the problems it poses for utilitarianism, see Finniss 1983: 86–93; Raz 1986: 321–68; George 1993: 88–90.

5. I do not believe that the preferences of the dead are always without moral weight. What happens after our death can affect how well our life went, and our desire for certain things after our death can be an important focus for our activities in life. Indeed, if the preferences of the dead did not sometimes have moral weight, it would be impossible to make sense of the way we treat wills. See the discussion in Lomasky (1987: 212–21), Hanser (1990), and Feinberg (1980: 173–6). On the 'experience requirement' more generally, see Scanlon (1991: 22–3), Larmore (1987: 48–9), Lomasky (1987: 231–3), Griffin (1986: 13–23), Parfit (1984: 149–53).

6. For a detailed exploration of this problem, see the essays in Elster and Roemer 1991.

7. Political theories which are concerned with the distribution of resources, without determining the effect these resources have on each person's welfare, may seem an exception to this general claim. But, as I will discuss in Chapter 3, this is a misleading perception, and even resource-based theories must have some theory of people's 'essential interests, most comprehensively construed' (Dworkin 1983: 24).

8. It is not clear whether utilitarianism can in fact limit itself to the basic structure of society, or to political decision-making. Even if utilitarianism applies in the first instance to political decisions or social institutions, and not to the personal conduct of individuals, one of the decisions governments face is to determine the legitimate scope of private attachments. If people are not maximizing utility in their private lives, then reorganizing the basic structure so as to leave less room for private life could increase utility. If comprehensive moral utilitarianism cannot accommodate our sense of the value of personal attachments, then political utilitarianism will have no reason to preserve a robust private realm. In any event, the predominance of utilitarianism in political philosophy stems mostly from the belief that it is the only coherent or systematic moral philosophy (Rawls 1971: pp. vii–viii), and so the motivation

for political utilitarianism is reduced if comprehensive moral utilitarianism can be shown to be indefensible.

9. The U-agent is often described as an 'act utilitarian', because he acts directly on the basis of utility calculations. But this is misleading in so far as 'act utilitarian' is commonly contrasted with 'rule utilitarian'. What defines the U-agent is that he uses utility maximization *directly* as a decision-procedure, and, as we will see, he could do this while focusing on rules rather than acts. The distinction between direct and indirect utilitarianism cuts across the distinction between act- and rule-utilitarianism (Railton 1984: 156–7). The first contrast is whether the principle of utility maximization is viewed as a decision-procedure or a standard of rightness, not whether the principle of utility maximization (as either a standard of rightness or a decision-procedure) applies to acts or rules.

10. For other influential statements of this 'alienation' objection, see Kagan 1989: 1–2; Railton 1984; Jackson 1991.

11. Two of the most influential recent defences of rule-utilitarianism are Harsanyi 1985 and Hardin 1988; cf. Ball 1990. There are in fact several different versions of rule-utilitarianism, each with its own strengths and weaknesses. For a helpful overview, see Scarre 1996: 122–32.

12. Bailey, for example, argues that while allowing Roman-style games may increase utility on a rule-utilitarian view, it is suboptimal, in the sense that we could do even better if we socialized people to get pleasure in other ways that do not involve harming others (Bailey 1997: 21, 144–5). In other words, Roman games are good, from a utilitarian point of view, in the sense of increasing overall utility compared to the status quo, but we could do even better, and so utilitarians should prefer some alternative. Bailey thinks this argument helps bring utilitarianism in line with our everyday intuitions. In reality, however, most people think the Roman games were evil, rather than merely suboptimal, and that whatever pleasure they gave to spectators should be accorded no moral weight. For Bailey, as for Hare (1982: 30) and Smart (1973: 25–6), there is no basis in utilitarianism for excluding such illegitimate preferences from the calculus.

13. For discussions of Government House utilitarianism, see Wolff 1996*a*: 131; Goodin 1995: ch. 4; Bailey 1997: 26, 152–3.

14. Unlike the rule utilitarian, who views promises as ingenious devices to maximize utility, the indirect utilitarian views our *beliefs about promises* as ingenious devices for maximizing utility. But people do not, and arguably cannot, view their moral beliefs this way (Smith 1988).

15. In defence of the teleological interpretation, Parfit gives the following sort of hypothetical example: imagine a woman who can choose whether to delay her pregnancy. If she gets pregnant now, the child will have a life worth living, but will not have a very happy life. If she delays the pregnancy for two months, the resulting child will have a happy and fulfilling life. Parfit argues that most people would view it as immoral not to delay the pregnancy, unless there were some urgent reason for proceeding immediately. Yet this judgement cannot be explained on the equal consideration interpretation, since no one would be harmed or wronged by not delaying the pregnancy. (The child resulting from the immediate pregnancy is not harmed by being born, since he prefers to be alive than never to have been born; the potential child of the delayed pregnancy is not harmed, because she does not exist.) So if it is morally wrong not to delay the pregnancy, as Parfit thinks, then it must be because we have an obligation to increase the overall amount of utility in the world, an obligation independent of our obligation to treat particular people with equal concern or respect (Parfit 1984: 358–61). And if the woman has an obligation to increase the overall amount of utility in the world by conceiving the happier child, then why wouldn't all of us have an obligation to increase overall utility by bringing additional children into the world? A couple who only wanted one child should instead have two or more children, even if this reduces the average utility of themselves

and their first child, so long as the utility of each additional child outweighs the loss to the existing family members.

I will let readers judge for themselves whether this is a plausible argument for teleological utilitarianism. In so far as people think the woman should delay her pregnancy, I suspect this is partly for prudential rather than moral reasons (i.e. we think she herself will be better off if she delays the pregnancy), and also partly because people wrongly think that it would be the same child born two months later, and hence that that particular child is harmed by being brought into the world 'too early'. Once we filter out the prudential reasons, and clarify that it would be an entirely different child born (i.e. the product of a different egg and sperm), then it is far from clear that there is anything morally wrong in consciously choosing to have a child who will be less happy than some other child one could have conceived.

16. Critics of utilitarianism also conflate the two versions. This is true, for example, of Rawls's claim that utilitarians ignore the separateness of persons. According to Rawls, utilitarians endorse the principle of maximizing utility because they generalize from the one-person case (it is rational for each individual to maximize her happiness), to the many-person case (it is rational for society to maximize its happiness). Rawls objects to this generalization because it treats society as if it were a single person, and so ignores the difference between trade-offs within one person's life and trade-offs across lives (Rawls 1971: 27; cf. Nozick 1974: 32–3; Gordon 1980: 40; Mackie 1984: 86–7). However, neither the egalitarian nor the teleological version of utilitarianism makes this generalization, and Rawls's claim rests on a conflation of the two. On this, see Kymlicka 1988*b*: 182–5; Freeman 1994; Cummins 1990; Quinn 1993.

17. This is only part of what equality requires, for there are obligations to those who are unable to help themselves, and Good Samaritan obligations to those who are in dire need. In these cases, we have obligations that are not tied to respecting people's rightful claims. I return to these issues in Chapter 9.

18. This shows why it is wrong to claim that Dworkin's egalitarian plateau is 'purely formal' or 'empty' since it is compatible with many different kinds of distributions (Hart 1979: 95–6; Goodin 1982: 89–90; Mapel 1989: 54; Larmore 1987: 62; Raz 1986: ch. 9). As Dworkin notes, this objection 'misunderstands the role of abstract concepts in political theory and debate' (Dworkin 1977: 368). The idea of treating people as equals is abstract, but not formal—on the contrary, it is a substantive ideal that excludes some theories (e.g. racist ones), and that sets a standard to which other theories aspire. The fact that an abstract concept needs to be interpreted, and that different theories interpret it in different ways, does not show that the concept is empty, or that one interpretation of that concept is as good as any other.

19. For example, Bailey defends a form of utilitarianism, but suggests that it is only appropriate for 'marginal' rather than 'global' analysis—i.e. we should not attempt to design institutions *de novo* on the basis of utilitarian principles, but should only invoke utilitarianism to make marginal changes to existing institutions if and when they start to fail due to changed circumstances (Bailey 1997: 15).

# 3

## LIBERAL EQUALITY

### 1. RAWLS'S PROJECT

#### (a) Intuitionism and utilitarianism

In the last chapter I argued that we need some or other theory of fair shares prior to the calculation of utility, for there are limits to the way individuals can be legitimately sacrificed for the benefit of others. If we are to treat people as equals, we must protect them in their possession of certain rights and liberties. But which rights and liberties?

Most of the political philosophy written in the last thirty years has been on this question. There are some people, as we have seen, who continue to defend utilitarianism. But there has been a marked shift away from the 'once widely-accepted old faith that some form of utilitarianism, if only we could discover the right form, *must* capture the essence of political morality' (Hart 1979: 77), and most contemporary political philosophers have hoped to find a systematic alternative to utilitarianism. John Rawls was one of the first to present such an alternative in his 1971 book *A Theory of Justice*. Many others had written about the counter-intuitive nature of utilitarianism. But Rawls starts his book by complaining that political theory was caught between two extremes: utilitarianism on the one side, and an incoherent jumble of ideas and principles on the other. Rawls calls this second option 'intuitionism', an approach which is little more than a series of anecdotes based on particular intuitions about particular issues.

Intuitionism is an unsatisfying alternative to utilitarianism, for while we do indeed have anti-utilitarian intuitions on particular issues, we also want an alternative theory which makes sense of those intuitions. We want a theory which shows why these particular examples elicit disapproval in us. But 'intuitionism' never gets beyond, or underneath, these initial intuitions to show how they are related, or to provide principles that underlie and give structure to them.

Rawls describes intuitionist theories as having two features: