Locke, J. 1690. *Two Treatises of Government*. London: Awnsham and John Churchill. Ed. Peter Laslett, Cambridge: Cambridge University Press, 1960.
Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rawls, J. 1985. Justice as fairness: political not metaphysical. *Philosophy and Public Affairs* 14: 223–51.
Rousseau, J.-J. 1762. *On the Social Contract*. Trans. Maurice Cranston, London: Penguin.
Scanlon, T.M. 1982. Contractualism and utilitarianism. In *Utilitarianism and Beyond*, ed. A. Sen and B. Williams (Cambridge: Cambridge University Press): 103–28.
Tyler, T.R. 1990. *Why People Obey the Law*. New Haven: Yale University Press.

modern utilitarianism. Utilitarianism is the view that one should do whatever will bring about the greatest amount of good. It was first clearly propounded in the eighteenth century by the philosopher Jeremy Bentham (1789). Leading figures in its subsequent development were John Stuart Mill (1863) and Henry Sidgwick (1874), both philosophers with a strong interest in economics. Throughout its history, economists have had a strong influence on the development of utilitarian thinking. Recently, work by the economist John Harsanyi (1953, 1955, 1977a) has been particularly influential. Important recent writings include Griffin (1986), the debate contained in Smart and Williams (1973) and the collection in Sen and Williams (1982).

The broad and imprecise statement that one should do whatever will bring about the greatest amount of good leaves plenty of scope for differing opinions among utilitarians, and there is no more precise formulation of their doctrine that all utilitarians would accept. Indeed, some would not accept even this broad statement. But it will be helpful to have a more detailed formulation as a benchmark for comparing alternative versions of utilitarianism. One formulation is the conjunction of the following principles.

Consequentialism. Of the acts that are available, one should do the one that will have the best consequences – that is to say, the one that will bring about the best state of affairs.
Personal good. The goodness of a state of affairs is its goodness for people.
Additivity. The goodness of a state of affairs for people is the total wellbeing of individual people.
Hedonism. A person's wellbeing is the preponderance of pleasure over pain in her life.

This is now an old-fashioned sort of utilitarianism, and few modern authors would accept all these principles without reservation. But they provide a convenient scheme for classifying the points that are at issue in modern utilitarian thinking. This essay will take the principles of consequentialism, personal good, additivity and hedonism in turn, and discuss some of the questions raised by each. Many of the new ideas that have entered utilitarianism recently are responses to objections that were made against earlier versions of the doctrine. So this tour of the issues will also provide a picture of the present standing of utilitarianism, and its unsolved problems.

CONSEQUENTIALISM: DIRECT AND INDIRECT UTILITARIANISM. 'Indirect' utilitarianism has become popular in recent decades. The most familiar version of it is rule utilitarianism (see Harsanyi 1977b; Hooker 1996), but other versions are possible. For instance, there is character utilitarianism and virtue utilitarianism. All versions accept the principles of ordinary, direct utilitarianism, except that they deny consequentialism: they deny you should do the act that will bring about the best consequences. Instead, they first of all make a claim about which rules you should live by, or what sort of character you should have, or what virtues you should espouse: you should live by the rules that will have the best consequences – that will most promote total wellbeing – or have the character that will have the best consequences, or espouse the virtues that will have the best consequences. Then these theories say you should do whatever act is prescribed by the rules you should live by, or issues from the character you should have, or from the virtues you should espouse.

The attraction of indirect utilitarianism is that it seems to offer a way of overcoming some of the objections that are commonly levelled against the direct version. Here are four of these objections and the indirect utilitarian's replies.

The first objection is that some implications of utilitarianism are intuitively simply wrong. For example, suppose the police are holding a person whom they know to be innocent, but whom the public believes to be guilty of a terrible crime. Suppose there will be riots if this person is not punished, leading to many deaths and great destruction. Direct utilitarianism will dictate that the innocent person should be punished, to avert the greater harm that will be caused by the riots. But intuition suggests it is wrong to punish an innocent person, even if there will be beneficial consequences, because it is unjust. On the other hand, a rule utilitarian can argue that on balance it best promotes total wellbeing to adopt rules of justice rather than rules of expediency, even if this occasionally leads to destructive riots. A just society is happier than an unjust one. So rule utilitarianism can support rules of justice, and agree with the intuition that it is wrong to punish the innocent person on this occasion.

A second objection is that even by a utilitarian's own criterion, people must live by some rules. Otherwise society would collapse, and that would be bad for total wellbeing. For instance, people must normally keep promises. If they did not, no one would be able to trust anyone else, and most social interaction would become impossible. Direct utilitarianism says one should keep a promise only if keeping it will contribute more to total wellbeing than breaking it will. But if you break a promise whenever it turns out better for the total wellbeing to do so, your promises cannot be relied on. If people commonly followed the prescription of direct utilitarianism, the whole institution of promising would fail, and this would be very bad for total wellbeing. On the other hand, adopting the rule of keeping your promises (except perhaps when the consequences will be extremely bad) may well have better consequences than adopting some other rule about promises. If so, rule utilitarianism will tell you to adopt this rule, and keep your promises, and that will be good for total wellbeing.

A third objection is that direct utilitarianism requires people to adopt a cold impartiality towards others, because it requires them to give equal weight to each person's well-being. But in a happy human society, people are inevitably partial. They care more for their loved ones and their neighbours than for other people. To require impartiality would make the world a much less happy place. It would damage the total of wellbeing, so direct utilitarianism would once more fail by its own criterion. On the other hand, indirect utilitarianism in the form of character utilitarianism would say that people ought to cultivate warm and loving characters, which make them partial towards people close to them, because having this sort of character best promotes total wellbeing.

A fourth objection is known as the 'demandingness' objection (see Kagan 1989). It says that, in asking us always to do what will bring about the most good, utilitarianism is implausibly demanding. For example, those of us who live in rich countries could do more good by sending most of our wealth to feed the world's poor than by keeping most of it for ourselves. So direct utilitarianism says this is what we should do. But the objectors find it implausible that morality really demands so big a sacrifice from us. One response to their objection is to adopt a version of rule utilitarianism. This version says we should act according to the rules that would have the best consequences if they were generally adopted. If people who live in rich countries generally adopted the rule of sending, say, 30% of their wealth to the poor, perhaps that would have better consequences than other rules, such as the rule of sending 70%. So according to this version, 30% is the proportion each of us should send. The fraction that would be most beneficial were it adopted as a general rule is no doubt not actually 30%, but whatever it is, it will definitely be less than the fraction of a person's wealth that would do the most good when considered on its own. So this version of rule utilitarianism is definitely less demanding than direct utilitarianism.

Some of these objections arise from a misunderstanding of direct utilitarianism. Few direct utilitarians would apply their theory only to overt acts such as keeping or breaking a promise. They will apply it to all acts, including the act of adopting a rule or undertaking some character-forming activity. For instance, a direct utilitarian would think you should adopt the rule of keeping your promises, if adopting this rule will contribute more to total wellbeing than failing to adopt it would. Indeed many direct utilitarians would apply utilitarianism to character traits and other dispositions, as well as to acts. So a direct utilitarian might think you should have the disposition of partiality towards your loved ones, if having this disposition better contributes to total wellbeing than other dispositions you might have, such as cool impartiality.

Indeed, direct and indirect utilitarians generally agree about the rules you should adopt and the dispositions you should have. They differ in what they think about the rightness of the acts that issue from these rules or dispositions. Suppose you ought to adopt the rule of keeping promises, and you have indeed adopted it. Now suppose that, following this rule, you keep a promise on some particular occasion. Rule utilitarians think you are automatically right to do so, because you are following a rule you ought to have adopted. Direct utilitarians think you are right only if keeping this particular promise better promotes total wellbeing than breaking it would have done. They believe that following a rule you ought to have adopted may sometimes lead you to do things you ought not to do. Similarly, a disposition you ought to have may lead you to do something you ought not to do. Intuitively, this seems to happen often. Surely, people ought to care more for their loved ones than for the general good, but sometimes this partiality, which they ought to have, leads them to do things they ought not to do. For instance, they may promote a relative in her job when she does not deserve it.

The issue between direct and indirect utilitarianism is over the truth of intuitions like this. Intuition is not always on the side of direct utilitarianism. In the example of punishing an innocent person, which I mentioned above, intuition suggests that, not only ought we to adopt rules of justice, but on each particular occasion we ought to follow them, even if the result is to reduce total wellbeing. This conflicts with direct utilitarianism. So the debate between direct and indirect utilitarianism continues.

CONSEQUENTIALISM: MORAL SUBJECTS. Utilitarianism is addressed to all of us. It tells each of us to pursue the total wellbeing of everyone. On the other hand, many economists assume that actually each of us will pursue only our own wellbeing – a view known as 'psychological egoism'. If psychological egoism is true, utilitarianism tells us to do something that actually we will not do, and this seems to make it a pointless doctrine, at best. But in fact many authors have found utilitarianism compatible with psychological egoism. Indeed, the founding figure Jeremy Bentham himself seems to have been a psychological egoist. He was not very interested in personal morality, but in the institutions and activities of government, and particularly in the law. He treated utilitarianism as a theory of public rather than private morality; he thought the law should be designed to bring about the greatest total of wellbeing. Similarly, utilitarianism has been used by some economists as a theory of public morality. For example, some modern economists have adopted it as a principle for designing the tax system (see, for instance, Mirrlees 1971, 1982). This public role is its typical application in law and economics. It restricts the scope of consequentialism: it says only that the government or social institutions should do what will have the best consequences, not people in general. This is one way of dealing with the demandingness objection too. As a personal morality, utilitarianism may be implausibly demanding, but it may nevertheless be a credible public morality.

The view of utilitarianism as a public morality does raise a question about the behaviour of legislators and public officials. These people are supposed to execute the public morality. But they are only people, so according to psychological egoism they will pursue only their own wellbeing. How can they be influenced in their public acts by the total wellbeing of everyone, as public utilitarianism requires them to be? It seems that public utilitarianism may require public officials to be utilitarian in their individual acts, so it may indeed conflict with psychological egoism.

PERSONAL GOOD: POPULATION. The principle of personal good claims that the goodness of a state of affairs is its goodness for people. Nearly all modern utilitarians would add in the good of animals as well. That is not now controversial, and when in this essay I speak of people's good, I implicitly include the good of animals with it.

However, there is a great deal of controversy over issues that involve the creation and destruction of people (or animals). The principle of personal good is intended to express a concern for people's wellbeing, which is one of the hallmarks of utilitarianism. Utilitarians are concerned to make people better off: to improve their wellbeing. But this concern can easily find itself misdirected when we come to evaluate acts that cause births or deaths, and it turns out that the principle of personal good does not express it accurately.

To see why, think of some putative act of policy that will cause the population to be bigger than it would otherwise have been (a change in the tax laws for families, for instance). To keep things simple, suppose the policy will add some new people to the population but make no difference to the wellbeing of existing people. To decide whether the policy is a good one according to the principle of personal good, we have to compare how good the world will be for people if the policy is implemented with how good it will be if the policy is not implemented. This depends on how the good of different people is aggregated together.

Additivity is one principle of aggregation, and let us first try out that one. Then the goodness of the world for people is the total of people's wellbeing. This total will be increased by adding new people, provided the wellbeing of the new people is positive. So our principles including additivity imply that the policy is a good one if this proviso is satisfied. But no one is made better off by this policy; the existing people are simply left as well off as they were anyway. So this conclusion does not properly reflect the utilitarian concern for making people better off. Adding people to the world does not meet this concern. It increases total wellbeing, but not by making anyone better off.

At first it may seem the problem is in the additive principle of aggregation. This thought seems to have been the stimulus for a new version of utilitarianism that has been dominant among utilitarian economists for the last fifty years. It is known as 'average utilitarianism', and it has a different aggregation principle. It takes the goodness of the world to be the average, rather than the total, of people's wellbeing. The idea is that the number of people who exist does not matter, what matters is the quality of life of those who do exist. Average utilitarianism is intended to express the utilitarian concern for making people better off. But actually it fails in this intention. Think about some policy that leaves all existing people exactly as well off as before, but adds some new people who are better off than the average. Adding these people will increase average wellbeing, so it is favoured by average utilitarianism. Nevertheless, it benefits no one; it does not satisfy the concern for making people better off. Indeed, one version of average utilitarianism has an opposite implication. It implies it is a good thing to kill people whose wellbeing is below the average, because this increases average wellbeing. Of course, this conclusion is absurd, and better versions of average utilitarianism are not committed to it. (They value the average of people's lifetime wellbeing: their wellbeing aggregated over their lives as a whole. Since killing a person reduces her lifetime wellbeing, it inevitably reduces the average.)

The real source of the problem is not in the aggregation principle at all. However we choose to aggregate wellbeing across people, there will always be some way of increasing the aggregate by adding people to the population, without benefiting existing people at all. Consequently, no aggregation principle can properly capture the concern to make people better off. The principle of personal good tries to express this concern, but whatever aggregation principle it is coupled with, it fails to do it properly. In recent years there has been a flood of literature that tries to accommodate changes of population within utilitarianism in a way that truly reflects the concern for people's wellbeing, or that at least reaches an acceptable compromise. (As a small sample, see Bayles 1976, Blackorby, Bossert and Donaldson 1997, Dasgupta 1994, Narveson 1967 and Parfit 1984, Part IV.) It has proved extraordinarily difficult to find principles that do not lead to paradoxes or implausible conclusions. This is one of the most hotly disputed areas of utilitarian thinking.

ADDITIVITY: THE VALUE OF EQUALITY. According to the principle of additivity, all that matters is the total of people's wellbeing. How that wellbeing is distributed amongst people does not matter at all. So utilitarianism gives no value to equality, at least not directly.

Surprisingly perhaps, in its early days utilitarianism had a reputation as an egalitarian doctrine. This was not entirely undeserved. One of Bentham's slogans was 'Each to count for one and none for more than one'. That is to say, utilitarians value the wellbeing of everyone in the society equally, whether they are lords or peasants. This is certainly a sort of egalitarianism. But valuing everyone's wellbeing equally is not the same as valuing equality in everyone's wellbeing. If it turns out that the way to maximize the total wellbeing is for a few people to be very well off indeed, and others very miserable, then this unequal arrangement is what utilitarianism recommends.

This inegalitarian potential within utilitarianism was recognized long ago. For one, the utilitarian economist Francis Edgeworth (1881) pointed it out, and embraced the inegalitarian conclusion. But for a long time other utilitarian economists – notably Alfred Marshall ([1890] 1920) and A.C. Pigou (1920) – insisted their doctrine did in fact favour equality, despite initial appearances. They relied on the assumption of diminishing marginal benefit. It is plausible to assume that the richer you are the greater is your wellbeing. But it is also plausible to assume that as you get richer, each addition to your wealth increases your wellbeing by less and less: additions have diminishing marginal benefit. If this is so, these economists argued, then any addition to a nation's wealth will add more to total wellbeing if it goes to poorer people rather than to richer ones. So the utilitarian aim of maximizing total wellbeing is best promoted by diverting wealth from the rich towards the poor.

This is a bad argument. It relies on a further hidden

assumption. Diminishing marginal benefit is an assumption about the benefits wealth brings to a single person; it says nothing about comparative benefits to different people. On its own, therefore, it cannot imply anything about the merits of redistributing wealth between people. The argument's hidden assumption is that the relation between a person's wealth and her wellbeing is the same for everyone. This is plainly false. No doubt some people are more 'susceptible' than others to the benefits of wealth, as Edgeworth put it: they derive more benefit than other people do from any given amount of wealth. The utilitarian aim is then best promoted by concentrating wealth disproportionately on to more susceptible people, rather than by distributing it equally. The egalitarian attitude of Marshall and Pigou results from a false assumption.

The inegalitarian implications of utilitarianism are by now well recognized. John Rawls expressed this complaint against it when he said that utilitarianism 'adopt[s] for society as a whole the principle of rational choice for one man' (Rawls 1971: 26–7). His point is that a single person might reasonably adopt the additive principle over her life, but not society as a whole over its members. A single person might reasonably sacrifice some wellbeing at one time in her life for the sake of a greater gain in wellbeing at another time; the greater gain more than compensates her for the lesser loss. Utilitarianism treats relations between people similarly: one person may be asked to make a sacrifice of wellbeing for the sake of a greater gain to someone else. Rawls thinks this is unjustified, because one person cannot be compensated for a loss by a gain to someone else.

ADDITIVITY: MODIFIED UTILITARIANISM. Rawls's own response was to reject utilitarianism wholesale in favour of a radically different theory of justice. A less radical response is to modify utilitarianism by adopting an aggregation principle that differs from the additive one in a way that gives some value to equality. This idea has been popular among many welfare economists (e.g. see Atkinson and Stiglitz 1980: 340). The idea is that we should value the wellbeing of badly off people more than the wellbeing of well-off people. Indeed, wellbeing itself has diminishing marginal value; the better off a person is, the less valuable it is to bring her more wellbeing. The value of wellbeing is not a linear function of wellbeing, but a strictly concave function. We do not assess the goodness of a state of affairs just by adding up people's wellbeing. Instead, we first transform each person's wellbeing by a strictly concave function that represents the value of her wellbeing. Then we add up the resulting transforms across people. The additive principle is:

$$G = g_1 + g_2 + g_3 + \ldots \qquad (1)$$

where G is the goodness of a state of affairs for people, and g_1, g_2, g_3 and so on are the people's wellbeings. The new modified principle is:

$$G = f(g_1) + f(g_2) + f(g_3) + \ldots \qquad (2)$$

where f is the strictly concave transformation that represents the value of a person's wellbeing.

In the philosophical literature, this idea is often called 'the priority view' or 'giving priority to the worse off' (e.g. Seanlon 1978; Temkin 1993). The additively separable form of (2) shows it is not strictly egalitarian, because it values each person's wellbeing independently of everyone else's. It does not compare one person's wellbeing with another person's, as an egalitarian would. Nevertheless, it does have the egalitarian consequence that, given some total of wellbeing, the more equally this wellbeing is distributed the better. The priority view is not strictly egalitarianism, but it has this egalitarian consequence whilst preserving many of the features of utilitarianism.

ADDITIVITY: HARSANYI'S DEFENCE. On the other hand, recent years have also seen a forceful new defence of the additive principle as a component of utilitarianism, particularly by John Harsanyi. In the 1950s Harsanyi developed no less than two new arguments for additivity. Both are founded on a theory of rationality that is not itself a moral theory: the standard theory of rational decision in the face of uncertainty, known as 'expected-utility theory'. Expected-utility theory says that a rational person, faced with a choice, chooses the option that has the greatest expectation of a quantity called 'utility'. Controversially, Harsanyi takes utility to be a measure of how well off a person is: utility measures wellbeing, that is to say.

Harsanyi's first argument (Harsanyi 1953) is of a type that was later taken up by Rawls to reach a different conclusion. It uses the idea of an 'original position'. Harsanyi imagines a person who is free to choose among a range of different societies to live in, each having the same population. She can choose any society, but she must do so behind a 'veil of ignorance': she does not know which position in the society she will occupy. She will have a chance of being the best-off person, and the same chance of being the worst-off person, or anywhere in between. This is a choice under uncertainty, because the subject has to choose a society whilst uncertain what the result will be for herself. So expected-utility theory can be applied. Under Harsanyi's interpretation of the theory, the person will, if rational, choose the society where her expectation of wellbeing is greatest. This is the society with the greatest average wellbeing. Since the populations are all the same, it will also have the greatest total wellbeing. Because this is the society a rational person would choose from behind a veil of ignorance, Harsanyi claims it is the best society. So he derives the additive principle: the best society is the one with the greatest total wellbeing.

Harsanyi's second argument for additivity (Harsanyi 1955) also sets out from expected-utility theory. It works by means of a remarkable piece of mathematics that is too complicated to reproduce here. It is not transparent. Nevertheless, it is a genuine and significant argument and not merely sleight-of-hand. Harsanyi's arguments together give significant new strength to the additive principle of utilitarianism. But of course they are not conclusive, and leave plenty of room for debate. One important bone of contention is Harsanyi's interpretation of utility as a measure of wellbeing (see Sen 1977, Broome 1991, Weymark 1991).

HEDONISM: THE PREFERENCIST ALTERNATIVE. Bentham took the view – hedonism – that a person's wellbeing consists in having pleasure and avoiding pain. He seems to have assumed that pleasure and pain were quantities that could be added and subtracted, and were comparable between people. To most modern authors these assumptions seem implausible. They believe there are more good and bad things in life than pleasure and pain. For example, there are achievement and failure, being loved and being hated, and so on. These things seem valuable in themselves, apart from the pleasure and pain they may lead to. Moreover, it seems implausible that pleasure and pain are simple, interpersonally comparable quantities.

If there are other good and bad things besides pleasure and pain, the question arises of how they are to be weighed against each other in determining a person's overall wellbeing. Many modern authors are subjectivist about this: they believe that what is good for a person is what the person herself values. And they often take her preferences to indicate what she values. So we arrive at:

Preferencism. One state of affairs is better for a person than another if and only if the person prefers it to the other.

Actually, preferencism in this form has few defenders. Most authors recognize that a person's preferences may be a poor indication of what she values, because they may be badly formed, perhaps hastily or imprudently, or on the basis of inaccurate information. So most authors rely on preferences that are idealized in some way, rather than on people's actual preferences (e.g., Brandt 1979). For instance, they may claim that one state of affairs is better for a person than another if and only if the person would prefer it if she were rational and well informed. Idealized preferencism is a modern alternative to hedonism. These two are by no means the only accounts of wellbeing that are available to utilitarians (see, e.g., Sen 1979; Griffin 1986), but they represent two broad schools amongst utilitarians: those who value the having of good experiences and those who value the satisfaction of desires.

Although preferencism is a radically different view from hedonism, economists sometimes fail to distinguish the two properly. The word 'satisfaction' can cause confusion. According to preferencism, what is good for a person is to have her preferences satisfied. It is easy to suppose that having her preferences satisfied will give a person a feeling of satisfaction. This feeling is a sort of pleasure. So it is easy to suppose that preferencism identifies a person's wellbeing with this sort of pleasure. But that is a bad mistake. The satisfaction of preferences is a quite different matter from feelings of satisfaction. Satisfying your preferences does not necessarily give you a feeling of satisfaction. For example, you may prefer that your children should prosper after you are dead, rather than endure a miserable life. If this preference is satisfied, you will feel no satisfaction, since you will be dead.

Historically, preferencism seems to have entered utilitarian thinking through the work of economists. Like utilitarianism, economic theory had Benthamite sources. W.S. Jevons (1871) was particularly responsible for importing Bentham's ideas into economics, and he expressed his economic theory in terms of pleasure and pain. But in the first decades of the twentieth century economists learnt to express themselves in terms of preferences. This led to a neater economic theory. Moreover, it was held to be sounder on epistemic grounds, because preferences were claimed to be observable, whereas pleasure and pain were not (see Robbins 1932). When economists turned to utilitarian moral theory it was natural for them to speak in terms of preferences there too.

HEDONISM: PROBLEMS OF PREFERENCISM. However, preferencism (idealized or not) is not enough for utilitarian purposes. Preferences can only determine whether one state of affairs is better or worse for a person than another. They do not determine how much better or worse. But utilitarianism needs to add up wellbeing across people, so it needs a quantitative – 'cardinal' – scale of wellbeing, and moreover one that is comparable between people. Something must be added to preferencism to achieve this, and the obvious thing to add is a notion of intensity of preference.

Expected-utility theory can supply a notion of intensity. Utility as defined within expected-utility theory is cardinal, and can plausibly be taken as a cardinal measure of intensity of preference. If intensity of preference is in turn taken to determine a quantitative scale of wellbeing, utility will be a cardinal measure of wellbeing. This measure is particularly agreeable to the spirit of preferencism, because expected-utility theory derives a person's utility entirely from her preferences. So utility inherits the epistemic privilege that is claimed for preferences: it is observable. Nevertheless, it remains controversial whether utility does truly measure a person's wellbeing (see Sen 1977; Broome 1991; Weymark 1991), and its epistemic advantage seems irrelevant to that question.

It is controversial whether utility measures wellbeing for a single person, but a bigger stumbling-block is comparisons between people. If preferencism is to be successfully adopted into utilitarianism, intensities of preferences must be comparable between people. And if the spirit of preferencism is to survive, the comparisons must themselves be based on preferences. But people's preferences seem to give us no basis for comparing intensities between people. This single point has led, and still leads, a great many economists to abandon utilitarianism. They take preferencism for granted, and they believe a measure of wellbeing must be based on preferences. Preferences supply no basis for interpersonal comparisons of wellbeing. Yet utilitarianism requires these interpersonal comparisons. Therefore utilitarianism has to go. During the last half-century, much of utilitarianism has been squeezed out of large areas of welfare economics, leaving only this one small residue:

Pareto principle. One state of affairs is better than another if one person prefers the one to the other and no one prefers the other to the one.

The Pareto principle assumes preferencism but requires no comparisons of wellbeing between people. Since it offers only a sufficient condition for one state of affairs to be better than another, and since the condition is rarely satisfied, it gives us only a very weak and fairly useless moral theory. Still, it does seem to be the natural end of preferencism.

There have been some attempts to eonstruet interpersonal comparisons of wellbeing on a prefereneist basis. The most detailed is John Harsanyi's (1977a: eh 4). Harsanyi uses the notion of an 'extended preferenee'. Suppose you would prefer to be Antony rather than Cleopatra; that is an example of an extended preferenee. An extended preferenee is a preferenee between possessing all the personal features of one person, and living that person's life, and possessing all the personal features of someone else, and living that person's life. Harsanyi hopes to eonstruet interpersonal comparisons of wellbeing on the basis of extended preferences. His attempt remains eontroversial (see Broome 1998), and has so far made little impression on the body of utilitarian thinking.

CONCLUSION. Until the 1970s, utilitarianism held a dominating position in the praetieal moral philosophy of the English-speaking world. Since that time, it has had a serious rival in eontraetualism, an ethical theory that was relaunched into modern thinking in 1971 by John Rawls's *A Theory of Justice*. There were even reports of utilitarianism's imminent death. But utilitarianism is now in a vigorous and healthy state. It is responding to familiar objeetions. It is faeing up to new problems sueh as the ethics of population. It has revitalized its foundations with new arguments. It has radieally changed its eoneeption of human wellbeing. It remains a eredible moral theory.

JOHN BROOME

See also BENTHAM, JEREMY; DISTRIBUTIVE JUSTICE; HUME, DAVID; JUSTICE; MODERN CONTRACTARIANISM; PARETO OPTIMALITY; RULE-GUIDED BEHAVIOUR; SOCIAL JUSTICE.

Subject classification: 3a; 3b.

BIBLIOGRAPHY

Atkinson, A.B. and Stiglitz, J.E. 1980. *Lectures on Public Economics*. New York: McGraw-Hill.

Bayles, M.D. (ed.) 1976. *Ethics and Population*. Cambridge, MA: Schenkman Publishing Co.

Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1996.

Blackorby, C., Bossert, W. and Donaldson, D. 1997. Critieal level utilitarianism and the population-ethics dilemma. *Economics and Philosophy* 13: 197–230.

Brandt, R. 1979. *A Theory of the Good and the Right*. Oxford: Clarendon Press; New York: Oxford Press University Press.

Broome, J. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Broome, J. 1998. Extended preferenees. In *Preferences*, ed. C. Fehige, G. Meggle and U. Wessels, New York: W. de Gruyter.

Dasgupta, P. 1994. Savings and fertility: ethical issues. *Philosophy and Public Affairs* 23: 99–127.

Edgeworth, F.Y. 1881. *Mathematical Psychics*. London: Kegan Paul.

Griffin, J. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: Clarendon Press.

Harsanyi, J.C. 1953. Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy* 61: 434–5.

Harsanyi, J.C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63: 309–21.

Harsanyi, J.C. 1977a. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Harsanyi, J.C. 1977b. Rule utilitarianism and deeision theory. *Erkenntnis* 11: 25–53.

Hooker, B. 1996. Ross-style pluralism versus rule-eonsequentialism. *Mind* 105: 531–52.

Jevons, W.S. 1871. *The Theory of Political Economy*. London: Macmillan.

Kagan, S. 1989. *The Limits of Morality*. Oxford: Clarendon Press; New York: Oxford University Press.

Marshall, A. 1890. *Principles of Economics*. London: Maemillan, 8th edn, 1920.

Mill, J.S. 1863. *Utilitarianism*. In his *Collected Works*, Volume 10, Toronto: University of Toronto Press, 1969.

Mirrlees, J.A. 1971. An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38: 175–208.

Mirrlees, J.A. 1982. The economic uses of utilitarianism. In *Utilitarianism and Beyond*, ed. A.K. Sen and B. Williams, Cambridge: Cambridge University Press.

Narveson, J. 1967. Utilitarianism and new generations. *Mind* 76: 62–72.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Pigou, A.C. 1920. *The Economics of Welfare*. London: Macmillan.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Robbins, L. 1932. *An Essay on the Nature and Significaнce of Economic Science*. London: Macmillan.

Scanlon, T.M. 1978. Rights, goals and fairness. In *Public and Private Morality*, ed. S. Hampshire, Cambridge: Cambridge University Press.

Sen, A.K. 1977. Non-linear social welfare functions: a reply to Professor Harsanyi. In *Foundational Problems in the Special Sciences*, ed. R. Buts and J. Hintikka, Dordreeht: Reidel.

Sen, A.K. 1979. Utilitarianism and welfarism. *Journal of Philosophy* 76: 463–89.

Sen, A.K. and Williams, B.A.O. (eds). 1982. *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.

Sidgwick, H. 1874. *The Methods of Ethics*. London: Maemillan, 7th edn, 1907.

Smart, J.J.C. and Williams, B.A.O. 1973. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Temkin, L. 1993. *Inequality*. New York: Oxford University Press.

Weymark, J.A. 1991. A reconsideration of the Harsanyi–Sen debate on utilitarianism. In *Interpersonal Comparisons of Well-Being*, ed. J. Roemer and J. Elster, Cambridge: Cambridge University Press.

moral hazard. *See* INSURANCE LAW.

moral rights. *See* COPYRIGHT; DROIT DE SUITE.

most-favoured-customer clauses. Most-favoured-eustomer elauses (MFCCs) are contraetual terms that guarantee buyers the right to purehase on terms as favourable as those offered to any other buyer; similar arrangements can also be erafted to provide sueh proteetion for sellers. They have attraeted considerable attention among eeonomists in reeent years for their ability to help firms create credible commitments. Three types of eommitment have been studied fairly extensively. First, MFCCs may help firms commit to esehew price discounting to partieular customers, thereby faeilitating taeit collusion. For example, if a manufacturer of gasoline additives commits to sell to all buyers at the same priee, then he