An outline of a system of utilitarian ethics

J. J. C. SMART
I. Introductory

Such writers as J. S. Mill, H. Sidgwick and G. E. Moore, as a result of philosophical reflection, produced systems of normative ethics. Of recent years normative ethics has become distinguished from meta-ethics, which discusses the nature of ethical concepts. Indeed, as a result of the prevalence of 'non-cognitivist' theories of meta-ethics, for example those of C. L. Stevenson¹ and R. M. Hare,² normative ethics has fallen into some disrepute, at any rate as a philosophical discipline. For non-cognitivist theories of ethics imply that our ultimate ethical principles depend on our ultimate attitudes and preferences. Ultimate ethical principles therefore seem to lie within the fields of personal decision, persuasion, advice and propaganda, but not within the field of academic philosophy.

While it is true that some ultimate ethical disagreements may depend simply on differences of ultimate preference, and while also the non-ultimate disagreements depend on differences about empirical facts, about which the philosopher is not specially qualified to judge, it nevertheless seems to me to be important to prevent this trend towards ethical neutrality of philosophy from going too far. The meta-ethical philosopher may far too readily forget that ordinary ethical thinking is frequently muddled, or else mixed up with questionable metaphysical assumptions. In the clear light of philosophical analysis some ethical systems may well come to seem less attractive. Moreover, even if there can be clear-headed disagreement about ultimate moral preferences, it is no small task to present one or other of the resulting ethical systems in a consistent and lucid manner, and in such a way as to show how

¹ *Ethics and Language* (Yale University Press, New Haven, 1944).
common, and often specious, objections to them can be avoided.

It will be my object in the present study to state a system of ethics which is free from traditional and theological associations. This is that type of utilitarianism which R. B. Brandt has called 'act-utilitarianism'. Roughly speaking, act-utilitarianism is the view that the rightness or wrongness of an action depends only on the total goodness or badness of its consequences, i.e. on the effect of the action on the welfare of all human beings (or perhaps all sentient beings). The best sustained exposition of act-utilitarianism is, I think, that in Sidgwick's Methods of Ethics, but Sidgwick stated it within the framework of a cognitivist meta-ethics which supposed that the ultimate act-utilitarian principles could be known to be true by some sort of intellectual intuition. I reject Sidgwick's meta-ethics for familiar reasons, and for the purpose of this study will assume the truth of some such 'non-cognitivist' meta-ethical analysis as that of Hare's Language of Morals, or possibly that of D. H. Monro in his Empiricism and Ethics. (Monro's theory should perhaps be classed as subjectivist rather than as non-cognitivist. However I am inclined to think that in the present state of linguistic theory it is not possible to make a very sharp distinction between these two sorts of theory. For our present purposes the distinction is unimportant, because both sorts of theory imply that a man's ultimate ethical principles depend on his attitudes or feelings.) In adopting such a meta-ethics, I do, of course, renounce the attempt to


prove the act-utilitarian system. I shall be concerned with stating it in a form in which it may appear persuasive to some people, and to show how it may be defended against many of the objections which are frequently brought up against utilitarianism. Nevertheless I should like to indicate my opinion that the choice of conceptually clear and emotionally attractive systems of normative ethics which might be alternatives to it is not as wide as is sometimes thought.

In the first place, B. H. Medlin¹ has argued that it is impossible to state ethical egoism without either confusion or else a sort of pragmatic inconsistency. Secondly, some widespread ethical systems depend partly on metaphysical premisses, and can therefore be undermined by philosophical criticism of these metaphysical bases. I myself would be prepared to argue that this is the case with respect to so-called 'natural law' ethics, which depends on a quasi-Aristotelian metaphysics. Thirdly, any system of deontological ethics, that is any system which does not appeal to the consequences of our actions, but which appeals to conformity with certain rules of duty, is open to a persuasive type of objection which may well be found convincing by some of those people who have the welfare of humanity at heart. For though, conceivably, in most cases the dictates of a deontological ethics might coincide with those of human welfare and of an act-utilitarian ethics, there must be some possible cases in which the dictates of the system clash with those of human welfare, indeed in which the deontological principles prescribe actions which lead to avoidable human misery. In the most attractive forms of deontological ethics the conflict with utilitarianism is in consequence of some principle of 'justice' or 'fairness', and I shall revert to this issue later.² In other cases,

² See pp. 67-73 below.
however, the conflict can be traced to some sort of confusion, perhaps even to some sort of superstitious ‘rule worship’. There is *prima facie* a necessity for the deontologist to defend himself against the charge of heartlessness, in his apparently preferring abstract conformity to a rule to the prevention of avoidable human suffering. Of course some deontologists might claim that though it is logically possible that their principles might conflict with the utilitarian one, *in fact* such a conflict would never occur. It seems that if such a deontology did exist, the utilitarian need not be concerned to defend himself against it, since its practical consequences would not differ from those of utilitarianism. However all deontological systems which are known to me do seem to differ from utilitarianism not only in theory but also in practice.

Such a ‘persuasive’ objection to deontology is possible simply *because* we have assumed the truth of non-cognitivist (or possibly, subjectivist) meta-ethics. A cognitivist in meta-ethics of the type of Sir David Ross\(^1\) could resist any such appeal to the heart by saying that whether we like it or not his deontological principles can be *seen* to be true. That they might sometimes conflict with human happiness or welfare might seem to him to be more of sentimental than of philosophic concern. But if we strip off the cognitivist meta-ethics from Ross’s theory, then his deontology may come to look artificial and perhaps infected by a sort of ‘rule worship’. For example the obligation to keep promises seems to be too artificial, to smack too much of human social conventions, to do duty as an ultimate principle. On the other hand it is, as we shall see, harder to produce persuasive arguments against a restrained deontology which supplements the utilitarian principle by principles related to abstract justice and fair distribution. However, I am not attempting

to show that the utilitarian can have no philosophically clear-headed rivals, but am merely trying to suggest that it is harder than is commonly believed to produce clear-headed and acceptable deontological systems of ethics, and that the range of these is probably not so wide as to embrace some of the well-known ones, such as that of Sir David Ross.

In setting up a system of normative ethics the utilitarian must appeal to some ultimate attitudes which he holds in common with those people to whom he is addressing himself. The sentiment to which he appeals is generalized benevolence, that is, the disposition to seek happiness, or at any rate, in some sense or other, good consequences, for all mankind, or perhaps for all sentient beings. His audience may not initially be in agreement with the utilitarian position. For example, they may have a propensity to obey the rules of some traditional moral system into which they have been indoctrinated in youth. Nevertheless the utilitarian will have some hope of persuading the audience to agree with his system of normative ethics. As a utilitarian he can appeal to the sentiment of generalized benevolence, which is surely present in any group with whom it is profitable to discuss ethical questions. He may be able to convince some people that their previous disposition to accept non-utilitarian principles was due to conceptual confusions. He will not be able to convince everybody, no doubt, but that utilitarianism will not be accepted by everybody, or even by all philosophically clear-headed people, is not in itself an objection to it. It may well be that there is no ethical system which appeals to all people, or even to the same person in different moods. I shall revert to this matter later on.¹

To some extent then, I shall be trying to present Sidgwick in a modern dress. The axioms of utilitarianism are no longer the deliverances of intellectual intuition but the expressions

¹ See pp. 72–3 below.
of our ultimate attitudes or feelings. Deductions from these axioms nevertheless go through in very much the same way. In a discussion note commenting on the earlier edition of this monograph, Charles Landesman suggested\(^1\) that as a non-cognitivist I am not entitled to talk about the logical consequences of ethical principles. However it is not clear to me that this is an insuperable difficulty. For example, R. M. Hare\(^2\) and others have worked out theories of logical relations between imperative sentences, and even mere expressions of attitude can be said to be consistent or inconsistent with one another.

Thus ‘Boo to snakes’ is consistent with ‘Boo to reptiles’ and inconsistent with ‘Hurrah for reptiles’. Indeed there is no reason why a non-cognitivist should refuse to call ethical sentences ‘true’ or ‘false’. He can say ‘‘Smith is good’’ is true if and only if Smith is good.’ He can even say things like ‘Some of Buddha’s ethical sayings are true’, thus giving to understand that he would be in agreement with some of the attitudes expressed in Buddha’s sayings, even though he is not telling, and even may not know, which ones these are. I must concede, however, that there are difficulties (attested to by the word ‘would’ in the previous sentence) in giving a proper semantics on these lines. The semantics for ‘would’ gets us into talk about possible worlds, which are dubious entities. Again consider a sentence like ‘If it rains Smith’s action is right.’ A non-cognitivist would perhaps interpret this as expressing approval of Smith’s action in a possible world in which it is raining. However ethics, whether non-cognitivist or not, probably needs the notion of a possible world,\(^3\) dubious or not, since it is concerned with alternative possible actions, and so in this respect the non-cognitivist

\(^2\) *The Language of Morals*.
may not really be worse off than the cognitivist. At any rate, I am assuming in this monograph that adequate non-cognitivist theories of meta-ethics exist.

2. Act-utilitarianism and rule-utilitarianism
The system of normative ethics which I am here concerned to defend is, as I have said earlier, act-utilitarianism. Act-utilitarianism is to be contrasted with rule-utilitarianism. Act-utilitarianism is the view that the rightness or wrongness of an action is to be judged by the consequences, good or bad, of the action itself. Rule-utilitarianism is the view that the rightness or wrongness of an action is to be judged by the goodness and badness of the consequences of a rule that everyone should perform the action in like circumstances. There are two sub-varieties of rule-utilitarianism according to whether one construes 'rule' here as 'actual rule' or 'possible rule'. With the former, one gets a view like that of S. E. Toulmin¹ and with the latter, one like Kant's.² That is, if it is permissible to interpret Kant's principle 'Act only on that maxim through which you can at the same time will that it should become a universal law' as 'Act only on that maxim which you as a humane and benevolent person would like to see established as a universal law.' Of course Kant would resist this appeal to human feeling, but it seems necessary in order to interpret his doctrine in a plausible way. A subtle version of the Kantian type of rule-utilitarianism is given by R. F. Harrod in his 'Utilitarianism Revised'.³

¹ *An Examination of the Place of Reason in Ethics* (Cambridge University Press, London, 1950).
³ *Mind* 45 (1936) 137–56.
I have argued elsewhere\(^1\) the objections to rule-utilitarianism as compared with act-utilitarianism.\(^2\) Briefly they boil down to the accusation of rule worship:\(^3\) the rule-utilitarian presumably advocates his principle because he is ultimately concerned with human happiness: why then should he advocate abiding by a rule when he knows that it will not in the present case be most beneficial to abide by it? The reply that in most cases it is most beneficial to abide by the rule seems irrelevant. And so is the reply that it would be better that everybody should abide by the rule than that nobody should. This is to suppose that the only alternative to ‘everybody does A’ is ‘no one does A’. But clearly we have the possibility ‘some people do A and some don’t’. Hence to refuse to break a generally beneficial rule in those cases in which it is not most beneficial to obey it seems irrational and to be a case of rule worship.

The type of utilitarianism which I shall advocate will, then, be act-utilitarianism, not rule-utilitarianism.

David Lyons has recently argued that rule-utilitarianism (by which, I think, he means the sort of rule-utilitarianism which I have called the Kantian one) collapses into act-utilitarianism.\(^4\) His reasons are briefly as follows. Suppose

\(^1\)In my article ‘Extreme and restricted utilitarianism’, *Philosophical Quarterly* 6 (1956) 344-54. This contains bad errors and a better version of the article will be found in Philippa Foot (ed.), *Theories of Ethics* (Oxford University Press, London, 1967), or Michael D. Bayles (ed.), *Contemporary Utilitarianism* (Doubleday, New York, 1968). In this article I used the terms ‘extreme’ and ‘restricted’ instead of Brandt’s more felicitous ‘act’ and ‘rule’ which I now prefer.


\(^4\)David Lyons, *The Forms and Limits of Utilitarianism* (Oxford University Press, London, 1965). Rather similar considerations have been put
that an exception to a rule R produces the best possible consequences. Then this is evidence that the rule R should be modified so as to allow this exception. Thus we get a new rule of the form ‘do R except in circumstances of the sort C’. That is, whatever would lead the act-utilitarian to break a rule would lead the Kantian rule-utilitarian to modify the rule. Thus an adequate rule-utilitarianism would be extensionally equivalent to act-utilitarianism.

Lyons is particularly interested in what he calls ‘threshold effects’. A difficulty for rule-utilitarianism has often appeared to be that of rules like ‘do not walk on the grass’ or ‘do not fail to vote at an election’. In these cases it would seem that it is beneficial if some people, though not too many, break the rule. Lyons points out that we can distinguish the action of doing something (say, walking on the grass) after some largish number n other people have done it from the action of doing it when few or no people have done it. When these extra circumstances are written into the rule, Lyons holds that the rule will come to enjoin the same actions as would the act-utilitarian principle. However there seems to be one interesting sort of case which requires slightly different treatment. This is the sort of case in which not too many people must do action X, but each person must plan his action in ignorance of what the other person does. That is, what A does depends on what B does, and what B does depends on what A does. Situations possessing this sort of circularity will be discussed below, pp. 57-62.

I am inclined to think that an adequate rule-utilitarianism would not only be extensionally equivalent to the act-utilitarian principle (i.e. would enjoin the same set of actions

as it) but would in fact consist of one rule only, the act-utilitarian one: 'maximize probable benefit'. This is because any rule which can be formulated must be able to deal with an indefinite number of unforeseen types of contingency. No rule, short of the act-utilitarian one, can therefore be safely regarded as extensionally equivalent to the act-utilitarian principle unless it is that very principle itself. I therefore suggest that Lyons' type of consideration can be taken even further, and that rule-utilitarianism of the Kantian sort must collapse into act-utilitarianism in an even stronger way: it must become a 'one-rule' rule-utilitarianism which is identical to act-utilitarianism. In any case, whether this is correct or not, it is with the defence of act-utilitarianism, and not with rule-utilitarianism (supposing that there are viable forms of rule-utilitarianism which may be distinguished from act-utilitarianism) that this monograph is concerned. (Lyons himself rejects utilitarianism.)

3. Hedonistic and non-hedonistic utilitarianism

An act-utilitarian judges the rightness or wrongness of actions by the goodness and badness of their consequences. But is he to judge the goodness and badness of the consequences of an action solely by their pleasantness and unpleasantness? Bentham, who thought that quantity of pleasure being equal, the experience of playing pushpin was as good as that of reading poetry, could be classified as a hedonistic act-utilitarian. Moore, who believed that some


states of mind, such as those of acquiring knowledge, had intrinsic value quite independent of their pleasantness, can be called an ideal utilitarian. Mill seemed to occupy an intermediate position. He held that there are higher and lower pleasures. This seems to imply that pleasure is a necessary condition for goodness but that goodness depends on other qualities of experience than pleasantness and unpleasantness. I propose to call Mill a quasi-ideal utilitarian. For Mill, pleasantness functions like \( x \) in the algebraic product, \( x \times y \times z \). If \( x = 0 \) the product is zero. For Moore pleasantness functions more like \( x \) in \( (x + 1) \times y \times z \). If \( x = 0 \) the product need not be zero. Of course this is only a very rough analogy.

What Bentham, Mill and Moore are all agreed on is that the rightness of an action is to be judged solely by consequences, states of affairs brought about by the action. Of course we shall have to be careful here not to construe ‘state of affairs’ so widely that any ethical doctrine becomes utilitarian. For if we did so we would not be saying anything at all in advocating utilitarianism. If, for example, we allowed ‘the state of having just kept a promise’, then a deontologist who said we should keep promises simply because they are promises would be a utilitarian. And we do not wish to allow this.

According to the type of non-cognitivist (or subjectivist) ethics that I am assuming, the function of the words ‘ought’ and ‘good’ is primarily to express approval, or in other words, to commend. With ‘ought’ we commend actions. With ‘good’ we may commend all sorts of things, but here I am concerned with ‘good’ as used to commend states of affairs or consequences of actions. Suppose we could know with certainty the total consequences of two alternative actions \( A \) and \( B \), and suppose that \( A \) and \( B \) are the only possible actions open to us. Then in deciding whether we

ought to do \( A \) or \( B \), the act-utilitarian would ask whether the total consequences of \( A \) are better than those of \( B \), or vice versa, or whether the total consequences are equal. That is, he commends \( A \) rather than \( B \) if he thinks that the total consequences of \( A \) are better than those of \( B \). But to say ‘better’ is itself to commend. So the act-utilitarian has to do a double evaluation or piece of commending. First of all he has to evaluate consequences. Then on the basis of his evaluation of consequences he has to evaluate the actions \( A \) and \( B \) which would lead to these two sets of consequences. It is easy to fail to notice that this second evaluation is needed, but we can see that it is necessary if we remind ourselves of the following fact. This is that a non-utilitarian, say a philosopher of the type of Sir David Ross, might agree with us in the evaluation of the relative merits of the total sets of consequences of the actions \( A \) and \( B \) and yet disagree with us about whether we ought to do \( A \) or \( B \). He might agree with us in the evaluation of total consequences but disagree with us in the evaluation of possible actions. He might say: “The total consequences of \( A \) are better than the total consequences of \( B \), but it would be unjust to do \( A \), for you promised to do \( B \).”

My chief concern in this study is with the second type of evaluation: the evaluation of actions. The utilitarian addresses himself to people who very likely agree with him as to what consequences are good ones, but who disagree with him about the principle that what we ought to do is to produce the best consequences. For a reason, which will appear presently, the difference between ideal and hedonistic utilitarianism in most cases will not usually lead to a serious disagreement about what ought to be done in practice. In this section, however, I wish to clear the ground by saying something about the first type of evaluation, the evaluation of consequences. It is with respect to this evaluation that Bentham, Mill and Moore differ from one another.
Let us consider Mill’s contention that it is ‘better to be Socrates dissatisfied than a fool satisfied’.

Mill holds that pleasure is not to be our sole criterion for evaluating consequences: the state of mind of Socrates might be less pleasurable than that of the fool, but, according to Mill, Socrates would be happier than the fool.

It is necessary to observe, first of all, that a purely hedonistic utilitarian, like Bentham, might agree with Mill in preferring the experiences of discontented philosophers to those of contented fools. His preference for the philosopher’s state of mind, however, would not be an intrinsic one. He would say that the discontented philosopher is a useful agent in society and that the existence of Socrates is responsible for an improvement in the lot of humanity generally. Consider two brothers. One may be of a docile and easy temperament: he may lead a supremely contented and unambitious life, enjoying himself hugely. The other brother may be ambitious, may stretch his talents to the full, may strive for scientific success and academic honours, and may discover some invention or some remedy for disease or improvement in agriculture which will enable innumerable men of easy temperament to lead a contented life, whereas otherwise they would have been thwarted by poverty, disease or hunger. Or he may make some advance in pure science which will later have beneficial practical applications. Or, again, he may write poetry which will solace the leisure hours and stimulate the brains of practical men or scientists, thus indirectly leading to an improvement in society. That is, the pleasures of poetry or mathematics may be extrinsically valuable in a way in which those of pushpin or sun-bathing may not be. Though the poet or

---

mathematician may be discontented, society as a whole may be the more contented for his presence.

Again, a man who enjoys pushpin is likely eventually to become bored with it, whereas the man who enjoys poetry is likely to retain this interest throughout his life. Moreover the reading of poetry may develop imagination and sensitivity, and so as a result of his interest in poetry a man may be able to do more for the happiness of others than if he had played pushpin and let his brain deteriorate. In short, both for the man immediately concerned and for others, the pleasures of poetry are, to use Bentham's word, more secund than those of pushpin.

Perhaps, then, our preference for poetry over pushpin is not one of intrinsic value, but is merely one of extrinsic value. Perhaps strictly in itself and at a particular moment, a contented sheep is as good as a contented philosopher. However it is hard to agree to this. If we did we should have to agree that the human population ought ideally to be reduced by contraceptive methods and the sheep population more than correspondingly increased. Perhaps just so many humans should be left as could keep innumerable millions of placid sheep in contented idleness and immunity from depredations by ferocious animals. Indeed if a contented idiot is as good as a contented philosopher, and if a contented sheep is as good as a contented idiot, then a contented fish is as good as a contented sheep, and a contented beetle is as good as a contented fish. Where shall we stop?

Maybe we have gone wrong in talking of pleasure as though it were no more than contentment. Contentment consists roughly in relative absence of unsatisfied desires; pleasure is perhaps something more positive and consists in a balance between absence of unsatisfied desires and presence of satisfied desires. We might put the difference in this way: pure unconsciousness would be a limiting case of contentment, but not of pleasure. A stone has no unsatisfied desires,
but then it just has no desires. Nevertheless, this consideration will not resolve the disagreement between Bentham and Mill. No doubt a dog has as intense a desire to discover rats as the philosopher has to discover the mysteries of the universe. Mill would wish to say that the pleasures of the philosopher were more valuable intrinsically than those of the dog, however intense these last might be.

It appears, then, that many of us may well have a preference not only for enjoyment as such but for certain sorts of enjoyment. And this goes for many of the humane and beneficent readers whom I am addressing. I suspect that they too have an intrinsic preference for the more complex and intellectual pleasures. This is not surprising. We must not underrate the mere brute strength of a hard and fit human being: by any standards man is a large and strong animal. Nevertheless above all else man owes his survival to his superior intelligence. If man were not a species which was inclined above all else to think and strive, we should not be where we are now. No wonder that men have a liking for intelligence and complexity, and this may become increasingly so in future. Perhaps some people may feel that my remarks here are somewhat too complacent, in view of the liking of so many people for low-grade entertainments, such as certain popular television programmes. But even the most avid television addict probably enjoys solving practical problems connected with his car, his furniture, or his garden. However unintellectual he might be, he would certainly resent the suggestion that he should, if it were possible, change places with a contented sheep, or even a lively and happy dog. Nevertheless, when all is said and done, we must not disguise the fact that disagreements in ultimate attitude are possible between those who like Mill have, and those who like Bentham have not, an intrinsic preference for the 'higher' pleasures. However it is possible for two people to disagree about ultimate ends and yet agree
in practice about what ought to be done. It is worth while enquiring how much practical ethics is likely to be affected by the possibility of disagreement over the question of Socrates dissatisfied versus the fool satisfied.

‘Not very much’, one feels like saying at first. We noted that the most complex and intellectual pleasures are also the most fecund. Poetry elevates the mind, makes one more sensitive, and so harmonizes with various intellectual pursuits, some of which are of practical value. Delight in mathematics is even more obviously, on Benthamite views, a pleasure worth encouraging, for on the progress of mathematics depends the progress of mankind. Even the most hedonistic schoolmaster would prefer to see his boys enjoying poetry and mathematics rather than neglecting these arts for the pleasures of marbles or the tuckshop. Indeed many of the brutish pleasures not only lack fecundity but are actually the reverse of fecund. To enjoy food too much is to end up fat, unhealthy and without zest or vigour. To enjoy drink too much is even worse. In most circumstances of ordinary life the pure hedonist will agree in his practical recommendations with the quasi-ideal utilitarian.

This need not always be so. Some years ago two psychologists, Olds and Milner, carried out some experiments with rats. Through the skull of each rat they inserted an electrode. These electrodes penetrated to various regions of the brain. In the case of some of these regions the rat showed

behaviour characteristics of pleasure when a current was passed from the electrode, in others they seemed to show pain, and in others the stimulus seemed neutral. That a stimulus was pleasure-giving was shown by the fact that the rat would learn to pass the current himself by pressing a lever. He would neglect food and make straight for this lever and start stimulating himself. In some cases he would sit there pressing the lever every few seconds for hours on end. This calls up a pleasant picture of the voluptuary of the future, a bald-headed man with a number of electrodes protruding from his skull, one to give the physical pleasure of sex, one for that of eating, one for that of drinking, and so on. Now is this the sort of life that all our ethical planning should culminate in? A few hours’ work a week, automatic factories, comfort and security from disease, and hours spent at a switch, continually electrifying various regions of one’s brain? Surely not. Men were made for higher things, one can’t help wanting to say, even though one knows that men weren’t made for anything, but are the product of evolution by natural selection.

It might be said that the objection to continual sensual stimulation of the above sort is that though it would be pleasant in itself it would be infecund of future pleasures. This is often so with the ordinary sensual pleasures. Excessive indulgence in the physical pleasures of sex may possibly have a debilitating effect and may perhaps interfere with the deeper feelings of romantic love. But whether stimulation by the electrode method would have this weakening effect and whether it would impair the possibility of future pleasures of the same sort is another matter. For example, there would be no excessive secretion of hormones. The whole biochemical mechanism would, almost literally, be short-circuited. Maybe, however, a person who stimulated himself by the electrode method would find it so enjoyable that he would neglect all other pursuits. Maybe if everyone became
an electrode operator people would lose interest in everything else and the human race would die out.

Suppose, however, that the facts turned out otherwise: that a man could (and would) do his full share of work in the office or the factory and come back in the evening to a few hours contented electrode work, without bad after-effects. This would be his greatest pleasure, and the pleasure would be so great intrinsically and so easily repeatable that its lack of fecundity would not matter. Indeed perhaps by this time human arts, such as medicine, engineering, agriculture and architecture will have been brought to a pitch of perfection sufficient to enable most of the human race to spend most of its time electrode operating, without compensating pains of starvation, disease and squalor. Would this be a satisfactory state of society? Would this be the millennium towards which we have been striving? Surely the pure hedonist would have to say that it was.

It is time, therefore, that we had another look at the concept of happiness. Should we say that the electrode operator was really happy? This is a difficult question to be clear about, because the concept of happiness is a tricky one. But whether we should call the electrode operator ‘happy’ or not, there is no doubt (a) that he would be contented and (b) that he would be enjoying himself.

Perhaps a possible reluctance to call the electrode operator ‘happy’ might come from the following circumstance. The electrode operator might be perfectly contented, might perfectly enjoy his electrode operating, and might not be willing to exchange his lot for any other. And we ourselves, perhaps, once we became electrode operators too, could become perfectly contented and satisfied. But nevertheless, as we are now, we just do not want to become electrode operators. We want other things, perhaps to write a book or get into a cricket team. If someone said ‘from tomorrow onwards you are going to be forced to be an electrode operator’ we
should not be pleased. Maybe from tomorrow onwards, once the electrode work had started, we should be perfectly contented, but we are not contented now at the prospect. We are not satisfied at being told that we would be in a certain state from tomorrow onwards, even though we may know that from tomorrow onwards we should be perfectly satisfied. All this is psychologically possible. It is just the obverse of a situation which we often find. I remember an occasion on which I was suspended by cable car half-way up a precipitous mountain. As the cable car creaked upwards, apparently so flimsily held above the yawning chasm below, I fervently wished that I had never come in it. When I bought the ticket for the cable car I knew that I should shortly be wishing that I had never bought it. And yet I should have been annoyed if I had been refused it. Again, a man may be very anxious to catch a bus, so as to be in time for a dental appointment, and yet a few minutes later, while the drill is boring into his tooth, may wish that he had missed that bus. It is, contrariwise, perfectly possible that I should be annoyed today if told that from tomorrow onwards I should be an electrode addict, even though I knew that from tomorrow onwards I should be perfectly contented.

This, I think, explains part of our hesitancy about whether to call the electrode operator 'happy'. The notion of happiness ties up with that of contentment: to be fairly happy at least involves being fairly contented, though it involves something more as well. Though we should be contented when we became electrode operators, we are not contented now with the prospect that we should become electrode operators. Similarly if Socrates had become a fool he might thereafter have been perfectly contented. Nevertheless if beforehand he had been told that he would in the future become a fool he would have been even more dissatisfied than in fact he was. This is part of the trouble about the
dispute between Bentham and Mill. The case involves the possibility of (a) our being contented if we are in a certain state, and (b) our being contented at the prospect of being so contented. Normally situations in which we should be contented go along with our being contented at the prospect of our getting into such situations. In the case of the electrode operator and in that of Socrates and the fool we are pulled two ways at once.

Now to call a person 'happy' is to say more than that he is contented for most of the time, or even that he frequently enjoys himself and is rarely discontented or in pain. It is, I think, in part to express a favourable attitude to the idea of such a form of contentment and enjoyment. That is, for A to call B 'happy', A must be contented at the prospect of B being in his present state of mind and at the prospect of A himself, should the opportunity arise, enjoying that sort of state of mind. That is, 'happy' is a word which is mainly descriptive (tied to the concepts of contentment and enjoyment) but which is also partly evaluative. It is because Mill approves of the 'higher' pleasures, e.g. intellectual pleasures, so much more than he approves of the more simple and brutish pleasures, that, quite apart from consequences and side effects, he can pronounce the man who enjoys the pleasures of philosophical discourse as 'more happy' than the man who gets enjoyment from pushpin or beer drinking.

The word 'happy' is not wholly evaluative, for there would be something absurd, as opposed to merely unusual, in calling a man who was in pain, or who was not enjoying himself, or who hardly ever enjoyed himself, or who was in a more or less permanent state of intense dissatisfaction, a 'happy' man. For a man to be happy he must, as a minimal condition, be fairly contented and moderately enjoying himself for much of the time. Once this minimal condition is satisfied we can go on to evaluate various types of content-
ment and enjoyment and to grade them in terms of happiness. Happiness is, of course, a long-term concept in a way that enjoyment is not. We can talk of a man enjoying himself at a quarter past two precisely, but hardly of a man being happy at a quarter past two precisely. Similarly we can talk of it raining at a quarter past two precisely, but hardly about it being a wet climate at a quarter past two precisely. But happiness involves enjoyment at various times, just as a wet climate involves rain at various times.

To be enjoying oneself, Ryle once suggested, is to be doing what you want to be doing and not to be wanting to do anything else,\(^1\) or, more accurately, we might say that one enjoys oneself the more one wants to be doing what one is in fact doing and the less one wants to be doing anything else. A man will not enjoy a round of golf if (a) he does not particularly want to play golf, or (b) though he wants to play golf there is something else he wishes he were doing at the same time, such as buying the vegetables for his wife, filling in his income tax forms, or listening to a lecture on philosophy. Even sensual pleasures come under the same description. For example the pleasure of eating an ice-cream involves having a certain physical sensation, in a way in which the pleasure of golf or of symbolic logic does not, but the man who is enjoying an ice-cream can still be said to be doing what he wants to do (have a certain physical sensation) and not to be wanting to do anything else. If his mind is preoccupied with work or if he is conscious of a pressing engagement somewhere else, he will not enjoy the physical sensation, however intense it be, or will not enjoy it very much.

The hedonistic ideal would then appear to reduce to a state of affairs in which each person is enjoying himself. Since, as we noted, a dog may, as far as we can tell, enjoy chasing a rat as much as a philosopher or a mathematician

may enjoy solving a problem, we must, if we adopt the purely hedonistic position, defend the higher pleasures on account of their fecundity. And that might not turn out to be a workable defence in a world made safe for electrode operators.

To sum up so far, happiness is partly an evaluative concept, and so the utilitarian maxim 'You ought to maximize happiness' is doubly evaluative. There is the possibility of an ultimate disagreement between two utilitarians who differ over the question of pushpin versus poetry, or Socrates dissatisfied versus the fool satisfied. The case of the electrode operator shows that two utilitarians might come to advocate very different courses of actions if they differed about what constituted happiness, and this difference between them would be simply an ultimate difference in attitude. Some other possibilities of the 'science fiction' type will be mentioned briefly on pp. 66-7 below. So I do not wish to say that the difference in ultimate valuation between a hedonistic and a non-hedonistic utilitarian will never lead to difference in practice.

Leaving these more remote possibilities out of account, however, and considering the decisions we have to make at present, the question of whether the 'higher' pleasures should be preferred to the 'lower' ones does seem to be of slight practical importance. There are already perfectly good hedonistic arguments for poetry as against pushpin. As has been pointed out, the more complex pleasures are incomparably more fecund than the less complex ones: not only are they enjoyable in themselves but they are a means to further enjoyment. Still less, on the whole, do they lead to disillusionment, physical deterioration or social disharmony. The connoisseur of poetry may enjoy himself no more than the connoisseur of whisky, but he runs no danger of a headache on the following morning. Moreover the question of whether the general happiness would be increased by replac-
ing most of the human population by a bigger population of contented sheep and pigs is not one which by any stretch of the imagination could become a live issue. Even if we thought, on abstract grounds, that such a replacement would be desirable, we should not have the slightest chance of having our ideas generally adopted.

So much for the issue between Bentham and Mill. What about that between Mill and Moore? Could a pleasurable state of mind have no intrinsic value at all, or perhaps even a negative intrinsic value? Are there pleasurable states of mind towards which we have an unfavourable attitude, even though we disregard their consequences? In order to decide this question let us imagine a universe consisting of one sentient being only, who falsely believes that there are other sentient beings and that they are undergoing exquisite torment. So far from being distressed by the thought, he takes a great delight in these imagined sufferings. Is this better or worse than a universe containing no sentient being at all? Is it worse, again, than a universe containing only one sentient being with the same beliefs as before but who sorrows at the imagined tortures of his fellow creatures? I suggest, as against Moore, that the universe containing the deluded sadist is the preferable one. After all he is happy, and since there is no other sentient being, what harm can he do? Moore would nevertheless agree that the sadist was happy, and this shows how happiness, though partly an evaluative concept, is also partly not an evaluative concept.

It is difficult, I admit, not to feel an immediate repugnance at the thought of the deluded sadist. If throughout our childhood we have been given an electric shock whenever we had tasted cheese, then cheese would have become immediately distasteful to us. Our repugnance to the sadist arises, naturally enough, because in our universe sadists invariably do harm. If we lived in a universe in which by

---

An outline of a system of utilitarian ethics

universe containing two million happy beings, each neither more nor less happy than any in the first universe? Or would you, as a humane and sympathetic person, give a preference to the second universe? I myself cannot help feeling a preference for the second universe.¹ But if someone feels the other way I do not know how to argue with him. It looks as though we have yet another possibility of disagreement within a general utilitarian framework.

This type of disagreement might have practical relevance. It might be important in discussions of the ethics of birth control. This is not to say that the utilitarian who values total, rather than average, happiness may not have potent arguments in favour of birth control. But he will need more arguments to convince himself than will the other type of utilitarian.

In most cases the difference between the two types of utilitarianism will not lead to disagreement in practice. For in most cases the most effective way to increase the total happiness is to increase the average happiness, and vice versa.

5. Negative utilitarianism

Sir Karl Popper has suggested² that we should concern ourselves not so much with the maximization of happiness as with the minimization of suffering. By ‘suffering’ we must understand misery involving actual pain, not just unhappiness. For otherwise the doctrine becomes unclear. Suppose that we found a new university. We may hope that indirectly research will help to minimize pains, but that is not the only

¹ This does not mean that I approve of the present explosive increase in world population. A typical member of an over-populated planet is not equally happy with a typical member of a moderately populated planet.

reason why we found universities. We do so partly because we want the happiness of understanding the world. But producing the happiness of understanding could equally well be thought of as removing the unhappiness of ignorance.

Let us see what sort of utilitarian position we should develop if we made the minimization of misery our sole ultimate ethical principle. The doctrine of negative utilitarianism, that we should concern ourselves with the minimization of suffering rather than with the maximization of happiness, does seem to be a theoretically possible one. It does, however, have some very curious consequences, which have been pointed out by my brother, R. N. Smart. In virtue of these very curious consequences I doubt whether negative utilitarianism will commend itself to many people, though it is always possible that someone might feel so attracted by the principle that he would accept it in spite of its consequences. For example it is possible to argue that a negative utilitarian would have to be in favour of exterminating the human race. It seems likely that Popper is himself not a utilitarian, and so a fortiori not a negative utilitarian. For alongside the negative utilitarian principle he sets two principles, that we should tolerate the tolerant, and that we should resist tyranny. It is hard to see how these principles could be deduced from the negative utilitarian principle, for surely, as my brother has pointed out, on this principle we should approve of a tyrannical but benevolent world exploder. Such a tyrant would prevent infinite future misery.

Even though we may not be attracted to negative utilitarianism as an ultimate principle, we may concede that the injunction ‘worry about removing misery rather than about promoting happiness’ has a good deal to recommend

2 Popper, *The Open Society and its Enemies*. 
it as a subordinate rule of thumb. For in most cases we can do most for our fellow men by trying to remove their miseries. Moreover people will be less ready to agree on what goods they would like to see promoted than they will be to agree on what miseries should be avoided. Mill and Bentham might disagree on whether poetry should be preferred to pushpin, but they would agree that an occasional visit to the dentist is preferable to chronic toothache. While there are so many positive evils in the world there is plenty of scope for co-operative effort among men who may nevertheless disagree to some extent as to what constitute positive goods.

6. Rightness and wrongness of actions

I shall now state the act-utilitarian doctrine. Purely for simplicity of exposition I shall put it forward in a broadly hedonistic form. If anyone values states of mind such as knowledge independently of their pleasurableness he can make appropriate verbal alterations to convert it from hedonistic to ideal utilitarianism. And I shall not here take sides on the issue between hedonistic and quasi-ideal utilitarianism. I shall concern myself with the evaluation signified by ‘ought’ in ‘one ought to do that which will produce the best consequences’, and leave to one side the evaluation signified by the word ‘best’.

Let us say, then, that the only reason for performing an action $A$ rather than an alternative action $B$ is that doing $A$ will make mankind (or, perhaps, all sentient beings) happier than will doing $B$. (Here I put aside the consideration that in fact we can have only probable belief about the effects of our actions, and so our reason should be more precisely stated as that doing $A$ will produce more probable benefit than will doing $B$. For convenience of exposition I shelve this question of probability for a page or two.) This is so
simple and natural a doctrine that we can surely expect that many of my readers will have at least some propensity to agree. For I am talking, as I said earlier, to sympathetic and benevolent men, that is, to men who desire the happiness of mankind. Since they have a favourable attitude to the general happiness, surely they will have a tendency to submit to an ultimate moral principle which does no more than express this attitude. It is true that these men, being human, will also have purely selfish attitudes. Either these attitudes will be in harmony with the general happiness (in cases where everyone's looking after his own interests promotes the maximum general happiness) or they will not be in harmony with the general happiness, in which case they will largely cancel one another out, and so could not be made the basis of an interpersonal discussion anyway. It is possible, then, that many sympathetic and benevolent people depart from or fail to attain a utilitarian ethical principle only under the stress of tradition, of superstition, or of unsound philosophical reasoning. If this hypothesis should turn out to be correct, at least as far as these readers are concerned, then the utilitarian may contend that there is no need for him to defend his position directly, save by stating it in a consistent manner, and by showing that common objections to it are unsound. After all, it expresses an ultimate attitude, not a liking for something merely as a means to something else. Save for attempting to remove confusions and discredit superstitions which may get in the way of clear moral thinking, he cannot, of course, appeal to argument and must rest his hopes on the good feeling of his readers. If any reader is not a sympathetic and benevolent man, then of course it cannot be expected that he will have an ultimate pro-attitude to human happiness in general. Also some good-hearted readers may reject the utilitarian position because of certain considerations relating to justice. I postpone discussion of these until pp. 67-73.
The utilitarian’s ultimate moral principle, let it be remembered, expresses the sentiment not of altruism but of benevolence, the agent counting himself neither more nor less than any other person. Pure altruism cannot be made the basis of a universal moral discussion because it might lead different people to different and perhaps incompatible courses of action, even though the circumstances were identical. When two men each try to let the other through a door first a deadlock results. Altruism could hardly commend itself to those of a scientific, and hence universalistic, frame of mind. If you count in my calculations why should I not count in your calculations? And why should I pay more attention to my calculations than to yours? Of course we often tend to praise and honour altruism even more than generalized benevolence. This is because people too often err on the side of selfishness, and so altruism is a fault on the right side. If we can make a man try to be an altruist he may succeed as far as acquiring a generalized benevolence.

Suppose we could predict the future consequences of actions with certainty. Then it would be possible to say that the total future consequences of action $A$ are such-and-such and that the total future consequences of action $B$ are so-and-so. In order to help someone to decide whether to do $A$ or to do $B$ we could say to him: ‘Envisage the total consequences of $A$, and think them over carefully and imaginatively. Now envisage the total consequences of $B$, and think them over carefully. As a benevolent and humane man, and thinking of yourself just as one man among others, would you prefer the consequences of $A$ or those of $B$?’ That is, we are asking for a comparison of one (present and future) total situation with another (present and future) total situation. So far we are not asking for a summation or calculation of pleasures or happiness. We are asking only for a comparison of total situations. And it seems clear that we can frequently make such a comparison and say that one total situation is better.
than another. For example few people would not prefer a total situation in which a million people are well-fed, well-clothed, free of pain, doing interesting and enjoyable work, and enjoying the pleasures of conversation, study, business, art, humour, and so on, to a total situation where there are ten thousand such people only, or perhaps 999,999 such people plus one man with toothache, or neurotic, or shivering with cold. In general, we can sum things up by saying that if we are humane, kindly, benevolent people, we want as many people as possible now and in the future to be as happy as possible. Someone might object that we cannot envisage the total future situation, because this stretches into infinity. In reply to this we may say that it does not stretch into infinity, as all sentient life on earth will ultimately be extinguished, and furthermore we do not normally in practice need to consider very remote consequences, as these in the end approximate rapidly to zero like the furthermost ripples on a pond after a stone has been dropped into it.

But do the remote consequences of an action diminish to zero? Suppose that two people decide whether to have a child or remain childless. Let us suppose that they decide to have the child, and that they have a limitless succession of happy descendants. The remote consequences do not seem to get less. Not at any rate if these people are Adam and Eve. The difference would be between the end of the human race and a limitless accretion of human happiness, generation by generation. The Adam and Eve example shows that the 'ripples on the pond' postulate is not needed in every case for a rational utilitarian decision. If we had some reason for thinking that every generation would be more happy than not we would not (in the Adam and Eve sort of case) need to be worried that the remote consequences of our action would be in detail unknown. The necessity for the 'ripples in the pond' postulate comes from the fact that usually we do not know whether remote consequences will be good or
bad. Therefore we cannot know what to do unless we can assume that remote consequences can be left out of account. This can often be done. Thus if we consider two actual parents, instead of Adam and Eve, then they need not worry about thousands of years hence. Not, at least, if we assume that there will be ecological forces determining the future population of the world. If these parents do not have remote descendants, then other people will presumably have more than they would otherwise. And there is no reason to suppose that my descendents would be more or less happy than yours. We must note, then, that unless we are dealing with ‘all or nothing’ situations (such as the Adam and Eve one, or that of someone in a position to end human life altogether) we need some sort of ‘ripples in the pond’ postulate to make utilitarianism workable in practice. I do not know how to prove such a postulate, though it seems plausible enough. If it is not accepted, not only utilitarianism, but also deontological systems like that of Sir David Ross, who at least admits beneficence as one prima facie duty among the others, will be fatally affected.

Sometimes, of course, more needs to be said. For example one course of action may make some people very happy and leave the rest as they are or perhaps slightly less happy. Another course of action may make all men rather more happy than before but no one very happy. Which course of action makes mankind happier on the whole? Again, one course of action may make it highly probable that everyone will be made a little happier whereas another course of action may give us a much smaller probability that everyone will be made very much happier. In the third place, one course of action may make everyone happy in a pig-like way, whereas another course of action may make a few people happy in a highly complex and intellectual way.

It seems therefore that we have to weigh the maximizing of happiness against equitable distribution, to weigh prob-
abilities with happiness, and to weigh the intellectual and other qualities of states of mind with their pleasurableness. Are we not therefore driven back to the necessity of some calculus of happiness? Can we just say: "envisage two total situations and tell me which you prefer"? If this were possible, of course there would be no need to talk of summing happiness or of a calculus. All we should have to do would be to put total situations in an order of preference. Since this is not always possible there is a difficulty, to which I shall return shortly.

We have already considered the question of intellectual versus non-intellectual pleasures and activities. This is irrelevant to the present issue because there seems to be no reason why the ideal or quasi-ideal utilitarian cannot use the method of envisaging total situations just as much as the hedonistic utilitarian. It is just a matter of envisaging various alternative total situations, stretching out into the future, and saying which situation one prefers. The non-hedonistic utilitarian may evaluate the total situations differently from the hedonistic utilitarian, in which case there will be an ultimate ethical disagreement. This possibility of ultimate disagreement is always there, though we have given reasons for suspecting that it will not frequently lead to important disagreement in practice.

Let us now consider the question of equity. Suppose that we have the choice of sending four equally worthy and intelligent boys to a medium-grade public school or of leaving three in an adequate but uninspiring grammar school and sending one to Eton. (For sake of the example I am making the almost certainly incorrect assumption that Etonians are happier than other public-school boys and that these other public-school boys are happier than grammar-school boys.) Which course of action makes the most for the happiness of the four boys? Let us suppose that we can neglect complicating factors, such as that the superior
Etonian education might lead one boy to develop his talents so much that he will have an extraordinary influence on the well-being of mankind, or that the unequal treatment of the boys might cause jealousy and rift in the family. Let us suppose that the Etonian will be as happy as (we may hope) Etonians usually are, and similarly for the other boys, and let us suppose that remote effects can be neglected. Should we prefer the greater happiness of one boy to the moderate happiness of all four? Clearly one parent may prefer one total situation (one boy at Eton and three at the grammar school) while another may prefer the other total situation (all four at the medium-grade public school). Surely both parents have an equal claim to being sympathetic and benevolent, and yet their difference of opinion here is not founded on an empirical disagreement about facts. I suggest, however, that there are not in fact many cases in which such a disagreement could arise. Probably the parent who wished to send one son to Eton would draw the line at sending one son to Eton plus giving him expensive private tuition during the holidays plus giving his other sons no secondary education at all. It is only within rather small limits that this sort of disagreement about equity can arise. Furthermore the cases in which we can make one person very much happier without increasing general happiness are rare ones. The law of diminishing returns comes in here. So, in most practical cases, a disagreement about what should be done will be an empirical disagreement about what total situation is likely to be brought about by an action, and will not be a disagreement about which total situation is preferable. For example the inegalitarian parent might get the other to agree with him if he could convince him that there was a much higher probability of an Etonian benefiting the human race, such as by inventing a valuable drug or opening up the mineral riches of Antarctica, than there is of a non-Etonian doing so. (Once more I should like to say
that I do not myself take such a possibility very seriously!)
I must again stress that since disagreement about what causes
produce what effects is in practice so much the most impor-
tant sort of disagreement, to have intelligent moral dis-
cussion with a person we do not in fact need complete
agreement with him about ultimate ends: an approximate
agreement is sufficient.

Rawls¹ has suggested that we must maximize the general
happiness only if we do so in a fair way. An unfair way of
maximizing the general happiness would be to do so by a
method which involved making some people less happy
than they might be otherwise.² As against this suggestion a
utilitarian might make the following rhetorical objection:
if it is rational for me to choose the pain of a visit to the
dentist in order to prevent the pain of toothache, why is it
not rational of me to choose a pain for Jones, similar to that
of my visit to the dentist, if that is the only way in which I
can prevent a pain, equal to that of my toothache, for
Robinson? Such situations continually occur in war, in
mining, and in the fight against disease, when we may often
find ourselves in the position of having in the general interest
to inflict suffering on good and happy men. However it
must be conceded that these objections against fairness as an
ultimate principle must be rhetorical only, and that Rawls's
principle could perhaps be incorporated in a restrained
system of deontological ethics, which would avoid the
artificiality of the usual forms of deontology. There are in
any case plenty of good utilitarian reasons for adopting the
principle of fairness as an important, but not inviolable, rule
of thumb.

We must now deal with the difficulty about probability.
We have so far avoided the common objection to utilitar-
ianism that it involves the allegedly absurd notion of a

¹ 'Justice as fairness', Philosophical Review 67 (1958) 164–94.
² See especially p. 168 of Rawls's article.
summation or calculus of happiness or goodness. We have
done this by using the method of comparing total situations.
All we have to do is to envisage two or more total situations
and say which we prefer. A purely ordinal, not a quantita-
tive, judgement is all we require. However in taking this
position we have oversimplified the matter. Unfortunately
we cannot say with certainty what would be the various
total situations which could result from our actions. Worse
still, we cannot even assign rough probabilities to the total
situations as a whole. All we can do is to assign various
probabilities to the various possible effects of an action. For
example, one course of action may almost certainly lead to
a fairly good result next year together with a high prob-
ability of a slightly good result the year after, while another
action may give a very small probability of a moderately
good result the year after and a very small but not negligible
probability of a rather bad result the year after that. (I am
assuming that in both cases the still more remote results
become negligible or such as to cancel one another out.) If
we had to weight total situations with probabilities, this
would give us enough conceptual difficulty, but it now
appears that we have to go within total situations and weight
different elements within them according to different prob-
abilities. We seem to be driven back towards a calculus.
If it were possible to assign numerical probabilities to the
various effects of our actions we could devise a way of applying
the method of total situations. Suppose that we could
say that an action X would either give Smith the pleasure of
eating ice-cream with probability 4/5 or the pain of tooth-
ache with probability 1/5 and that it would give Jones the
pleasure of sympathy with probability 3/5 or the displeasure
of envy with probability 2/5 and that no other important
results (direct or indirect) would accrue. Suppose that the
only alternative action to X is Y and that this has no effect
on Smith but causes Jones to go to sleep with probability 3/5
or to go for a walk with probability \( \frac{2}{5} \) and that no other important results (direct or indirect) would accrue. Then we could say that the total situations we have to imagine and to compare are (a) (for \( X \)): four people (just like Smith) eating ice-cream plus one (just like Smith) with toothache plus three sympathetic people (just like Jones) plus two envious people (just like Jones), and (b) (for \( Y \)): three people (just like Jones) who are asleep plus two (just like Jones) going for a walk. In the example I have, for convenience, taken all probabilities to be multiples of \( \frac{1}{5} \). If they did not have common denominators we should have to make them such, by expressing them as multiples of a denominator which is the lowest common multiple of the original denominators.

However it is not usually possible to assign a numerical probability to a particular event. No doubt we could use actuarial tables to ascertain the probability that a friend of ours, who is of a certain age, a certain carefully specified medical history, and a certain occupation, will die within the next year. But can we give a numerical value to the probability that a new war will break out, that a proof of Fermat's last theorem will be found, or that our knowledge of genetical linkage in human chromosomes will be much improved in the next five years? Surely it is meaningless to talk of a numerical value for these probabilities, and it is probabilities of this sort with which we have to deal in our moral life.

When, however, we look at the way in which in fact we take some of our ordinary practical decisions we see that there is a sense in which most people think that we can weigh up probabilities and advantages. A man deciding whether to migrate to a tropical country may well say to himself, for example, that he can expect a pleasanter life for himself and his family in that country, unless there is a change in the system of government there, which is not
very likely, or unless one of his children catches an epidemic
disease, which is perhaps rather more likely, and so on, and
thinking over all these advantages and disadvantages and
probabilities and improbabilities he may come out with the
statement that on the whole it seems preferable for him to
go there or with the statement that on the whole it seems
preferable for him to stay at home.

If we are able to take account of probabilities in our
ordinary prudential decisions it seems idle to say that in the
field of ethics, the field of our universal and humane atti-
tudes, we cannot do the same thing, but must rely on some
dogmatic morality, in short on some set of rules or rigid
criteria. Maybe sometimes we just will be unable to say
whether we prefer for humanity an improbable great
d Advantage or a probable small advantage, and in these cases
perhaps we shall have to toss a penny to decide what to do.
Maybe we have not any precise methods for deciding what
to do, but then our imprecise methods must just serve their
turn. We need not on that account be driven into authori-
tarianism, dogmatism or romanticism.

So, at any rate, it appears at first sight. But if I cannot
say any more the utilitarian position as it is here presented
has a serious weakness. The suggested method of developing
normative ethics is to appeal to feelings, namely of benevo-
ience, and to reason, in the sense of conceptual clarification
and also of empirical enquiry, but not, as so many moralists
do, to what the ordinary man says or thinks. The ordinary
man is frequently irrational in his moral thinking. And if he
can be irrational about morals why cannot he be irrational
about probabilities? The fact that the ordinary man thinks
that he can weigh up probabilities in making prudential
decisions does not mean that there is really any sense in
what he is doing. What utilitarianism badly needs, in order
to make its theoretical foundations secure, is some method
according to which numerical probabilities, even approxi-
mate ones, could in theory, though not necessarily always in practice, be assigned to any imagined future event.

D. Davidson and P. Suppes have proposed a method whereby, at any rate in simplified situations, subjective probabilities can be given a numerical value. Their theory was to some extent anticipated in an essay by F. P. Ramsey, in which he tries to show how numbers can be assigned to probabilities in the sense of degrees of belief. This allows us to give a theory of rational, in the sense of self-consistent, utilitarian choice, but to make utilitarianism thoroughly satisfactory we need something more. We need a method of assigning numbers to objective, not subjective, probabilities. Perhaps one method might be to accept the Davidson-Suppes method of assigning subjective probabilities, and define objective probabilities as the subjective probabilities of an unbiased and far-sighted man. This, however, would require independent criteria for lack of bias and for farsightedness. I do not know how to do this, but I suspect, from the work that is at present being done on decision-making, that the situation may not be hopeless. But until we have an adequate theory of objective probability utilitarianism is not on a secure theoretical basis. Nor, for that matter, is ordinary prudence; nor are deontological systems of ethics, like that of Sir David Ross, which assign some weight to beneficence. And any system of deontological ethics implies some method of weighing up the claims of conflicting prima facie duties, for it is impossible that deontological rules of conduct should never conflict, and the

---

3 R. McNaughton's interesting article 'A metrical concept of happiness', Philosophy and Phenomenological Research 14 (1953–4) 171–83, does not enable us to propose a complete utilitarian calculus, because it neglects probability considerations.
rationale of this is perhaps even more insecure than is the theory of objective probability.

7. The place of rules in act-utilitarianism

According to the act-utilitarian, then, the rational way to decide what to do is to decide to perform that one of those alternative actions open to us (including the null-action, the doing of nothing) which is likely to maximize the probable happiness or well-being of humanity as a whole, or more accurately, of all sentient beings. The utilitarian position is here put forward as a criterion of rational choice. It is true that we may choose to habituate ourselves to behave in accordance with certain rules, such as to keep promises, in the belief that behaving in accordance with these rules is generally optimific, and in the knowledge that we most often just do not have time to work out individual pros and cons. When we act in such an habitual fashion we do not of course deliberate or make a choice. The act-utilitarian will, however, regard these rules as mere rules of thumb, and will use them only as rough guides. Normally he will act in accordance with them when he has no time for considering probable consequences or when the advantages of such a consideration of consequences are likely to be outweighed by the disadvantage of the waste of time involved. He acts in accordance with rules, in short, when there is no time to think, and since he does not think, the

1 In the first edition of this monograph I said 'which is likely to bring about the total situation now and in the future which is the best for the happiness or well-being of humanity as a whole, or more accurately, of all sentient beings'. This is inaccurate. To probably maximize the benefit is not the same as to maximize the probable benefit. This has been pointed out by David Braybrooke. See p. 35 of his article 'The choice between utilitarianisms', American Philosophical Quarterly 4 (1967) 28–38.
actions which he does habitually are not the outcome of moral thinking. When he has to think what to do, then there is a question of deliberation or choice, and it is precisely for such situations that the utilitarian criterion is intended.

It is, moreover, important to realize that there is no inconsistency whatever in an act-utilitarian’s schooling himself to act, in normal circumstances, habitually and in accordance with stereotyped rules. He knows that a man about to save a drowning person has no time to consider various possibilities, such as that the drowning person is a dangerous criminal who will cause death and destruction, or that he is suffering from a painful and incapacitating disease from which death would be a merciful release, or that various timid people, watching from the bank, will suffer a heart attack if they see anyone else in the water. No, he knows that it is almost always right to save a drowning man, and in he goes. Again, he knows that we would go mad if we went in detail into the probable consequences of keeping or not keeping every trivial promise: we will do most good and reserve our mental energies for more important matters if we simply habituate ourselves to keep promises in all normal situations. Moreover he may suspect that on some occasions personal bias may prevent him from reasoning in a correct utilitarian fashion. Suppose he is trying to decide between two jobs, one of which is more highly paid than the other, though he has given an informal promise that he will take the lesser paid one. He may well deceive himself by underestimating the effects of breaking the promise (in causing loss of confidence) and by overestimating the good he can do in the highly paid job. He may well feel that if he trusts to the accepted rules he is more likely to act in the way that an unbiased act-utilitarian would recommend than he would be if he tried to evaluate the consequences of his possible actions himself. Indeed Moore argued on
act-utilitarian grounds that one should never in concrete cases think as an act-utilitarian.¹

This, however, is surely to exaggerate both the usefulness of rules and the human mind's propensity to unconscious bias. Nevertheless, right or wrong, this attitude of Moore's has a rational basis and (though his argument from probability considerations is faulty in detail) is not the law worship of the rule-utilitarian, who would say that we ought to keep to a rule that is the most generally optimific, even though we knew that obeying it in this particular instance would have bad consequences.

Nor is this utilitarian doctrine incompatible, as M. A. Kaplan² has suggested it is, with a recognition of the importance of warm and spontaneous expressions of emotion. Consider a case in which a man sees that his wife is tired, and simply from a spontaneous feeling of affection for her he offers to wash the dishes. Does utilitarianism imply that he should have stopped to calculate the various consequences of his different possible courses of action? Certainly not. This would make married life a misery and the utilitarian knows very well as a rule of thumb that on occasions of this sort it is best to act spontaneously and without calculation. Moreover I have said that act-utilitarianism is meant to give a method of deciding what to do in those cases in which we do indeed decide what to do. On these occasions when we do not act as a result of deliberation and choice, that is, when we act spontaneously, no method of decision, whether utilitarian or non-utilitarian, comes into the matter. What

¹ Principia Ethica, p. 162.

² Morton A. Kaplan, 'Some problems of the extreme utilitarian position', Ethics 70 (1959-60) 228–32. This is a critique of my earlier article 'Extreme and restricted utilitarianism', Philosophical Quarterly 6 (1956) 344–54. He also puts forward a game theoretic argument against me, but this seems cogent only against an egoistic utilitarian. Kaplan continued the discussion in his interesting note 'Restricted utilitarianism', Ethics 71 (1960–1) 301–2.
does arise for the utilitarian is the question of whether or not he should consciously encourage in himself the tendency to certain types of spontaneous feeling. There are in fact very good utilitarian reasons why we should by all means cultivate in ourselves the tendency to certain types of warm and spontaneous feeling.

Though even the act-utilitarian may on occasion act habitually and in accordance with particular rules, his criterion is, as we have said, \textit{applied} in cases in which he does not act habitually but in which he deliberates and chooses what to do. Now the right action for an agent in given circumstances is, we have said, that action which produces better results than any alternative action. If two or more actions produce equally good results, and if these results are better than the results of any other action open to the agent, then there is no such thing as \textit{the} right action: there are two or more actions which are \textit{a} right action. However this is a very exceptional state of affairs, which may well never in fact occur, and so usually I will speak loosely of the action which is \textit{the} right one. We are now able to specify more clearly what is meant by ‘alternative action’ here. The fact that the utilitarian criterion is meant to apply in situations of deliberation and choice enables us to say that the class of alternative actions which we have in mind when we talk about an action having the best possible results is the class of actions which the agent could have performed if he had tried. For example, it would be better to bring a man back to life than to offer financial assistance to his dependants, but because it is technologically impossible to bring a man back to life, bringing the man back to life is not something we could do if we tried. On the other hand it may well be possible for us to give financial assistance to the dependants, and this then may be the right action. The right action is the action among those which we could do, i.e. those which we \textit{would} do if we chose to, which has the best possible results.
It is true that the general concept of action is wider than that of deliberate choice. Many actions are performed habitually and without deliberation. But the actions for whose rightness we as agents want a criterion are, in the nature of the case, those done thinkingly and deliberately. An action is at any rate that sort of human performance which it is appropriate to praise, blame, punish or reward, and since it is often appropriate to praise, blame, punish, or reward habitual performances, the concept of action cannot be identified with that of the outcome of deliberation and choice. With habitual actions the only question that arises for an agent is that of whether or not he should strengthen the habit or break himself of it. And individual acts of habit-strengthening or habit-breaking can themselves be deliberate.

The utilitarian criterion, then, is designed to help a person, who could do various things if he chose to do them, to decide which of these things he should do. His utilitarian deliberation is one of the causal antecedents of his action, and it would be pointless if it were not. The utilitarian view is therefore perfectly compatible with determinism. The only sense of 'he could have done otherwise' that we require is the sense 'he would have done otherwise if he had chosen'. Whether the utilitarian view necessitates complete metaphysical determinism is another matter. All that it requires is that deliberation should determine actions in the way that everyone knows it does anyway. If it is argued that any indeterminism in the universe entails that we can never know the outcome of our actions, we can reply that in normal cases these indeterminacies will be so numerous as approximately to cancel one another out, and anyway all that we require for rational action is that some consequences of our actions should be more probable than others, and this is something which no indeterminist is likely to deny.

The utilitarian may now conveniently make a terminological recommendation. Let us use the word 'rational' as a
term of commendation for that action which is, on the evidence available to the agent, *likely* to produce the best results, and to reserve the word ‘right’ as a term of commendation for the action which does *in fact* produce the best results. That is, let us say that what is rational is to try to perform the right action, to try to produce the best results. Or at least this formulation will do where there is an equal probability of achieving each possible set of results. If there is a very low probability of producing very good results, then it is natural to say that the rational agent would perhaps go for other more probable though not quite so good results. For a more accurate formulation we should have to weight the goodness of the results with their probabilities. However, neglecting this complication, we can say, roughly, that it is rational to perform the action which is on the available evidence the one which will produce the best results. This allows us to say, for example, that the agent did the right thing but irrationally (he was trying to do something else, or was trying to do this very thing but went about it unscientifically) and that he acted rationally but by bad luck did the wrong thing, because the things that seemed probable to him, for the best reasons, just did not happen.

Roughly, then: we shall use ‘right’ and ‘wrong’ to appraise choices on account of their actual success in promoting the general happiness, and we shall use ‘rational’ and ‘irrational’ to appraise them on account of their likely success. As was noted above (p. 42) ‘likely success’ must be interpreted in terms of maximizing the probable benefit, not in terms of probably maximizing the benefit. In effect, it is rational to do what you reasonably think to be right, and what will be right is what will maximize the probable benefit. We need, however, to make one qualification to this. A person may unreasonably believe what it would in fact be reasonable to believe. We shall still call such a person’s action irrational.
An outline of a system of utilitarian ethics

If the agent has been unscientific in his calculation of means–ends relationships he may decide that a certain course of action is probably best for human happiness, and it may indeed be so. When he performs this action we may still call his action irrational, because it was pure luck, not sound reasoning, that brought him to his conclusion.

‘Rational’ and ‘irrational’ and ‘right’ and ‘wrong’ so far have been introduced as terms of appraisal for chosen or deliberate actions only. There is no reason why we should not use the pair of terms ‘right’ and ‘wrong’ more widely so as to appraise even habitual actions. Nevertheless we shall not have much occasion to appraise actions that are not the outcome of choice. What we do need is a pair of terms of appraisal for agents and motives. I suggest that we use the terms ‘good’ and ‘bad’ for these purposes. A good agent is one who acts more nearly in a generally optimific way than does the average one. A bad agent is one who acts in a less optimific way than the average. A good motive is one which generally results in beneficent actions, and a bad motive is one which generally ends in maleficent actions. Clearly there is no inconsistency in saying that on a particular occasion a good man did a wrong action, that a bad man did a right action, that a right action was done from a bad motive, or that a wrong action was done from a good motive. Many specious arguments against utilitarianism come from obscuring these distinctions. Thus one may be got to admit that an action is ‘right’, meaning no more than that it is done from a good motive and is praiseworthy, and then it is pointed out that the action is not ‘right’ in the sense of being optimific. I do not wish to legislate as to how other people (particularly non-utilitarians) should use words like ‘right’ and ‘wrong’, but in the interests of clarity it is important for me to state how I propose to use them myself, and to try to keep the various distinctions clear.

It should be noted that in making this terminological
recommendation I am not trying to smuggle in valuations under the guise of definitions, as Ardon Lyon, in a review of the first edition of this monograph, has suggested that I have done. It is merely a recommendation to pre-empt the already evaluative words ‘rational’ and ‘irrational’ for one lot of commendatory or discommendatory jobs, the already evaluative words ‘right’ and ‘wrong’ for another lot of commendatory or discommendatory jobs, and the already evaluative words ‘good’ and ‘bad’ for yet another lot of commendatory or discommendatory jobs.

We can also use ‘good’ and ‘bad’ as terms of commendation or discommendation of actions themselves. In this case to commend or discommend an action is to commend or discommend the motive from which it sprang. This allows us to say that a man performed a bad action but that it was the right one, or that he performed a good action but that it was wrong. For example, a man near Berchtesgaden in 1938 might have jumped into a river and rescued a drowning man, only to find that it was Hitler. He would have done the wrong thing, for he would have saved the world a lot of trouble if he had left Hitler below the surface. On the other hand his motive, the desire to save life, would have been one which we approve of people having: in general, though not in this case, the desire to save life leads to acting rightly. It is worth our while to strengthen such a desire. Not only should we praise the action (thus expressing our approval of it) but we should perhaps even give the man a medal, thus encouraging others to emulate it. Indeed praise itself comes to have some of the social functions of medal giving: we come to like praise for its own sake, and are thus influenced by the possibility of being given it. Praising a person is thus an important action in itself – it has significant effects. A utilitarian must therefore learn to control his acts of praise and dispraise, thus perhaps concealing his approval

of an action when he thinks that the expression of such approval might have bad effects, and perhaps even praising actions of which he does not really approve. Consider, for example, the case of an act-utilitarian, fighting in a war, who succeeds in capturing the commander of an enemy submarine. Assuming that it is a just war and that the act-utilitarian is fighting on the right side, the very courage and ability of the submarine commander has a tendency which is the reverse of optimistic. Everything that the submarine commander has been doing was (in my proposed sense of the word) wrong. (I do not of course mean that he did anything wrong in the technological sense: presumably he knew how to manoeuvre his ship in the right way.) He has kept his boat cunningly concealed, when it would have been better for humanity if it had been a sitting duck, he has kept the morale of his crew high when it would have been better if they had been cowardly and inefficient, and has aimed his torpedoes with deadly effect so as to do the maximum harm. Nevertheless, once the enemy commander is captured, or even perhaps before he is captured, our act-utilitarian sailor does the right thing in praising the enemy commander, behaving chivalrously towards him, giving him honour and so on, for he is powerfully influencing his own men to aspire to similar professional courage and efficiency, to the ultimate benefit of mankind.

What I have said in the last paragraph about the occasional utility of praising harmful actions applies, I think, even when the utilitarian is speaking to other utilitarians. It applies even more when, as is more usually the case, the utilitarian is speaking to a predominantly non-utilitarian audience. To take an extreme case, suppose that the utilitarian is speaking to people who live in a society governed by a form of magical taboo ethics. He may consider that though on occasion keeping to the taboos does harm, on the whole the tendency of the taboo ethics is more beneficial than the sort
of moral anarchy into which these people might fall if their reverence for their taboos was weakened. While, therefore, he would recognize that the system of taboos which governed these people's conduct was markedly inferior to a utilitarian ethic, nevertheless he might also recognize that these people's cultural background was such that they could not easily be persuaded to adopt a utilitarian ethic. He will, therefore, on act-utilitarian grounds, distribute his praise and blame in such a way as to strengthen, not to weaken, the system of taboo.

In an ordinary society we do not find such an extreme situation. Many people can be got to adopt a utilitarian, or almost utilitarian, way of thought, but many cannot. We may consider whether it may not be better to throw our weight on the side of the prevailing traditional morality, rather than on the side of trying to improve it with the risk of weakening respect for morality altogether. Sometimes the answer to this question will be 'yes', and sometimes 'no'. As Sidgwick said:

The doctrine that Universal Happiness is the ultimate standard must not be understood to imply that Universal Benevolence is... always the best motive of action. For... it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are to be preferred on Utilitarian principles.

In general, we may note, it is always dangerous to influence a person contrary to his conviction of what is right. More harm may be done in weakening his regard for duty than would be saved by preventing the particular action in question. Furthermore, to quote Sidgwick again, "any particular existing moral rule, though not the ideally best even for such beings, as existing men under the existing circumstances,"

1Methods of Ethics, p. 413.
may yet be the best that they can be got to obey”.¹ We must also remember that some motives are likely to be present in excess rather than defect; in which case, however necessary they may be, it is not expedient to praise them. It is obviously useful to praise altruism, even though this is not pure generalized benevolence, the treating of oneself as neither more nor less important than anyone else, simply because most people err on the opposite side, from too much self-love and not enough altruism. It is, similarly, inexpedient to praise self-love, important though this is when it is kept in due proportion. In short, to quote Sidgwick once more, “in distributing our praise of human qualities, on utilitarian principles, we have to consider not primarily the usefulness of the quality, but the usefulness of the praise”.²

Most men, we must never forget, are not act-utilitarians, and do not use the words ‘good’ and ‘bad’, when applied to agents or to motives, quite in the way which has here been recommended. When a man says that another is wicked he may even be saying something of a partly metaphysical or superstitious connotation. He may be saying that there is something like a yellow stain on the other man’s soul. Of course he would not think this quite literally. If you asked him whether souls could be coloured, or whether yellow was a particularly abhorrent colour, he would of course laugh at you. His views about sin and wickedness may be left in comfortable obscurity. Nevertheless the things he does say may indeed entail something like the yellow stain view. ‘Wicked’ has thus come to have much more force than the utilitarian ‘likely to be very harmful’ or ‘probably a menace’. To stigmatize a man as wicked is not, as things are, just to make men wary of him, but to make him the object of a peculiar and very powerful abhorrence, over and above the

¹ Ibid. p. 469.
² Ibid. p. 428.
natural abhorrence one has from a dangerous natural object such as a typhoon or an octopus. And it may well be to the act-utilitarian’s advantage, qua act-utilitarian, to acquiesce in this way of talking when he is in the company of non-utilitarians. He himself will not believe in yellow stains in souls, or anything like it. Tout comprendre c’est tout pardonner; a man is the result of heredity and environment. Nevertheless the utilitarian may influence behaviour in the way he desires by using ‘wicked’ in a quasi-superstitious way. Similarly a man about to be boiled alive by cannibals may usefully say that an imminent eclipse is a sign of the gods’ displeasure at the proposed culinary activities. We have seen that in a completely utilitarian society the utility of praise of an agent’s motives does not always go along with the utility of the action. Still more may this be so in a non-utilitarian society.

I cannot stress too often the importance of Sidgwick’s distinction between the utility of an action and the utility of praise or blame of it, for many fallacious ‘refutations’ of utilitarianism depend for their plausibility on confusing the two things.

Thus A. N. Prior1 quotes the nursery rhyme:

For want of a nail
The shoe was lost;
For want of a shoe
The horse was lost;
For want of a horse
The rider was lost;
For want of a rider
The battle was lost;
For want of a battle
The kingdom was lost;
And all for the want
Of a horse-shoe nail.

So it was all the blacksmith's fault! But, says Prior, it is surely hard to place on the smith's shoulders the responsibility for the loss of the kingdom. This is no objection, however, to act-utilitarianism. The utilitarian could quite consistently say that it would be useless to blame the blacksmith, or at any rate to blame him more than for any other more or less trivial case of 'bad maintenance'. The blacksmith had no reason to believe that the fate of the kingdom would depend on one nail. If you blame him you may make him neurotic and in future even more horses may be badly shod.

Moreover, says Prior, the loss of the kingdom was just as much the fault of someone whose negligence led to there being one fewer cannon in the field. If it had not been for this other piece of negligence the blacksmith's negligence would not have mattered. Whose was the responsibility? The act-utilitarian will quite consistently reply that the notion of the responsibility is a piece of metaphysical nonsense and should be replaced by 'Whom would it be useful to blame?' And in the case of such a close battle, no doubt it would be useful to blame quite a lot of people though no one very much. Unlike, for example, the case where a battle was lost on account of the general getting drunk, where considerable blame of one particular person would clearly be useful.

"But wouldn't a man go mad if he really tried to take the whole responsibility of everything upon himself in this way?" asks Prior. Clearly he would. The blacksmith must not mortify himself with morbid thoughts about his carelessness. He must remember that his carelessness was of the sort that is usually trivial, and that a lot of other people were equally careless. The battle was just a very close thing. But this refusal to blame himself, or blame himself very much, is surely consistent with the recognition that his action was in fact very wrong, that much harm would have been pre-
vented if he had acted otherwise. Though if other people, e.g. the man whose fault it was that the extra cannon did not turn up, had acted differently, then the blacksmith’s action would have in fact not been very wrong, though it would have been no more and no less blameworthy. A very wrong action is usually very blameworthy, but on some occasions, like the present one, a very wrong action can be hardly blameworthy at all. This seems paradoxical at first, but paradox disappears when we remember Sidgwick’s distinction between the utility of an action and utility of praise of it.

The idea that a consistent utilitarian would go mad with worry about the various effects of his actions is perhaps closely connected with a curious argument against utilitarianism to be found in Baier’s book *The Moral Point of View.* Baier holds that (act-) utilitarianism must be rejected because it entails that we should never relax, that we should use up every available minute in good works, and we do not ordinarily think that this is so. The utilitarian has two effective replies. The first is that perhaps what we ordinarily think is false. Perhaps a rational investigation would lead us to the conclusion that we should relax much less than we do. The second reply is that act-utilitarian premisses do not entail that we should never relax. Maybe relaxing and doing few good works today increases threefold our capacity to do good works tomorrow. So relaxation and play can be defended even if we ignore, as we should not, their intrinsic pleasures.

I beg the reader, therefore, if ever he is impressed by any alleged refutation of act-utilitarianism, to bear in mind the distinction between the rightness or wrongness of an action and the goodness or badness of the agent, and Sidgwick’s correlative and most important distinction between the

---

utility of an action and the utility of praise or blame of it. The neglect of this distinction is one of the commonest causes of fallacious refutations of act-utilitarianism.

It is also necessary to remember that we are here considering utilitarianism as a normative system. The fact that it has consequences which conflict with some of our particular moral judgements need not be decisive against it. In science general principles must be tested by reference to particular facts of observation. In ethics we may well take the opposite attitude, and test our particular moral attitudes by reference to more general ones. The utilitarian can contend that since his principle rests on something so simple and natural as generalized benevolence it is more securely founded than our particular feelings, which may be subtly distorted by analogies with similar looking (but in reality totally different) types of case, and by all sorts of hangovers from traditional and uncritical ethical thinking.

If, of course, act-utilitarianism were put forward as a descriptive systematization of how ordinary men, or even we ourselves in our unreflective and uncritical moments, actually think about ethics, then of course it is easy to refute and I have no wish to defend it. Similarly again if it is put forward not as a descriptive theory but as an explanatory one.

John Plamenatz, in his *English Utilitarians*, seems to hold that utilitarianism "is destroyed and no part of it left standing".1 This is apparently on the ground that the utilitarian explanation of social institutions will not work: that we cannot explain various institutions as having come about because they lead to the maximum happiness. In this monograph I am not concerned with what our moral customs and institutions in fact are, and still less am I concerned with the question of why they are as they in fact are. I am concerned with a certain view about what they ought to be. The correct-

ness of an ethical doctrine, when it is interpreted as recommendatory, is quite independent of its truth when it is interpreted as descriptive and of its truth when it is interpreted as explanatory. In fact it is precisely because a doctrine is false as description and as explanation that it becomes important as a possible recommendation.

8. Simple application of game-theory technique

So far I hope that I have shown that act-utilitarianism, as a normative theory of ethics, is not so simple-minded a doctrine as its critics seem to suppose, and that it escapes some of the usual refutations. I wish now to analyse a type of situation which has in the past proved difficult for the act-utilitarian to handle, but for which some very simple techniques of the theory of games seem to provide the solution.

R. B. Brandt\(^1\) considers the case of a utilitarian in wartime England, and it is supposed that there is a governmental request that a maximum temperature of 50° F. should be maintained in homes, so as to conserve gas and electricity. A utilitarian Frenchman who is resident in England might conceivably reason as follows: “It is very unlikely that the vast majority of Englishmen will not comply with this request. But it will do no harm at all if a few people, such as myself, live in a temperature of 70° F. And it will do these few people a lot of good for their comfort. Therefore the general happiness will be increased by my using enough electricity and gas to make myself comfortable.” The Frenchman thus decides to use the electricity and gas. Of course in practice such a decision might not make the Frenchman happier. If he was a decent person, normally brought up, he would feel very considerable twinges of

\(^1\) Ethical Theory, p. 389.
conscience. But suppose the Frenchman is an absolutely single-minded out and out utilitarian. What then?

The act-utilitarian will have to agree that if the Frenchman's behaviour could be kept secret then he ought in this case to use the electricity and gas. But the Frenchman should also agree that he should be condemned and punished if he were found out. There would indeed, as Brandt points out, be a horrible outcry if it became known that members of the Cabinet, who were aware of the willingness of most people to sacrifice and thus knew that electricity and gas were in reasonably good supply, ignored their own regulation. In this case, too, the utilitarian calculation would indeed be different if we assumed that the behaviour of the members of the Cabinet would leak out. Moreover the utilitarian would hold that in this case there would be good utilitarian reasons (especially in a generally non-utilitarian society) for condemning the Cabinet. We must recollect the distinction between utility of an action and utility of praise or blame of it. However, independently of this last point, we may agree that Brandt has produced a case in which the utilitarian is likely to conflict with common sense ethics. The utilitarian, to be consistent, must be willing to say, "So much the worse for common sense ethics!"

Brandt further objects that if everyone followed the Frenchman's reasoning disastrous results would follow. This objection fails to recognize that the Frenchman would have used as an empirical premise in his calculation the proposition that very few people would be likely to reason as he does. They would very likely be adherents of a traditional, non-utilitarian morality.

How would the Frenchman reason if he were living in a society composed entirely of convinced and rational act-utilitarians like himself? He is in the situation of not knowing how to plan his actions unless he has premisses about what other people will do, and each of them will not know how
to plan his actions unless he knows what the rest of the people (including the Frenchman) will do. There is a circularity in the situation which cries out for the technique of game theory.

There are three types of possibility: (a) he can decide to obey the government's request; (b) he can decide not to obey the government's request; (c) he can decide to give himself a certain probability of not obeying the government's request, e.g. by deciding to throw dice and disobey the government's request if and only if he got a certain number of successive sixes.

To decide to do something of type (c) is to adopt what in game theory is called 'a mixed strategy'. On plausible assumptions it would turn out that the best result would be attained if each member of the act-utilitarian society were to give himself a very small probability $p$ of disobeying the government's request. In practice $p$ is very difficult to calculate, and since it is likely to be very small, in practice the act-utilitarian will adopt alternative (a). Indeed if the trouble of calculating $p$ outweighed the probable benefit of adopting the mixed strategy, and we took this into account, we should have to plump for alternative (a) anyway.

Let us see how this probability $p$ could be calculated. Even if the matter is of little practical importance it is of interest for the theoretical understanding of ethics.

Let $m$ be the number of people in the community. Let $f(n)$ be the national damage done by exactly $n$ people disobeying the government's request; it will be an increasing function of $n$. Now if each member of the community gives himself a probability $p$ of disobeying the edict it is easy to determine, as functions of $p$, the probabilities $p_1, p_2, \ldots p_m$ of exactly $1, 2, \ldots m$ persons respectively disobeying the edict. Let $a$ be the personal benefit to each person of disobeying the edict. I am, of course, supposing what is
perhaps a fiction, that numerical values can be given to \( f(n) \) and to \( a \). Then if \( V \) is the total probable benefit to the community we have

\[
V = p_1(a-f(1)) + p_2(2a-f(2)) + p_3(3a-f(3)) + \ldots + p_m(ma-f(m)).
\]

If we know the function \( f(n) \) we can calculate the value of \( p \) for which \( \frac{dV}{dp} = 0 \). This will give the value of \( p \) which maximizes \( V \).

As I said, the matter is of theoretical rather than practical importance, as in the sort of case which I have in mind \( p \) will be so near to zero that the act-utilitarian would not bother to calculate but would just obey the government’s request. No doubt special examples of moral decision could be devised in which a not too small value of \( p \) would be obtained. This type of reasoning seems to be important more for the theoretical insight it affords than for its potentiality for practical guidance.¹

It might be thought that this symmetrical solution by means of mixed strategies implies some sort of rule-utilitarianism.² For will a group of act-utilitarians have any empirical basis for assuming that they will all adopt a symmetrical solution to the problem? Of course if David Lyons is right that rule-utilitarianism and act-utilitarianism collapse into one another, the problem disappears. However, as I remarked on p. 11 above, I cannot see how to apply Lyons’ argument

¹ The adoption of a mixed strategy would seem to provide the solution (in theory) to the garden watering example in A. K. Stout’s article, ‘But suppose everyone did the same’, *Australasian Journal of Philosophy* 32 (1954) 1–29.

to the sort of situation in which what one person ought to
do depends on what others ought to do and vice versa. This
sort of situation requires special treatment.

The clue lies in the notion of a convention, which has been
elucidated in an important book by David K. Lewis.\(^1\)
Lewis, in turn, makes use of Thomas C. Schelling’s study of
‘co-ordination games’,\(^2\) which suggests that two agents can
co-ordinate their activities without rules. For example, two
parachutists, who have been dropped in enemy country and
need to rendezvous, will both make their way to a bridge
when this is the only salient feature on the map. The act-
utilitarian will have to take this propensity to co-ordinate
behaviour as an empirical fact about human beings which
each will legitimately take into account when planning his
strategy. Lewis shows that the notion of a convention is
prior to that of a rule, and so I think that a reliance by the
act-utilitarian on conventions need not turn him into a
rule-utilitarian or even into a Kantian. Lewis has made a
remarkable analysis of one type of alleged objection to act-
utilitarianism, making use of his theory of convention, in an
article ‘Utilitarianism and truthfulness’.\(^3\)

Even if the solution to the present difficulty were rule-
utilitarianism it would be a rule-utilitarianism (or perhaps
Kantianism) which would be markedly different from those
which have generally been put forward, since it would be
applicable only in those situations in which all the agents are
utilitarians. My sort of utilitarian will normally think that
he ought to act when he is in a predominantly non-utilitarian
society in a way which is different from the way in which

\(^1\) David K. Lewis, *Convention* (Harvard University Press, Cambridge,

\(^2\) Thomas C. Schelling, *The Strategy of Conflict* (Harvard University

\(^3\) *Australasian Journal of Philosophy* 50 (1972) 17–19. This is in reply to an
argument by D. H. Hodgson, *Consequences of Utilitarianism* (Oxford
he ought to act when he is in a utilitarian society. Furthermore, even in the case of a society of like-minded utilitarians, the mixed strategy solution makes it importantly different from the usual 'all or none' varieties of rule-utilitarianism.

9. Utilitarianism and the future

The chief persuasive argument in favour of utilitarianism has been that the dictates of any deontological ethics will always, on some occasions, lead to the existence of misery that could, on utilitarian principles, have been prevented. Thus if the deontologist says that promises always should be kept (or even if, like Ross, he says that there is a *prima facie* duty to keep them) we may confront him with a situation like the following, the well-known 'desert island promise': I have promised a dying man on a desert island, from which subsequently I alone am rescued, to give his hoard of gold to the South Australian Jockey Club. On my return I give it to the Royal Adelaide Hospital, which, we may suppose, badly needs it for a new X-ray machine. Could anybody deny that I had done rightly without being open to the charge of heartlessness? (Remember that the promise was known only to me, and so my action will not in this case weaken the general confidence in the social institution of promising.) Think of the persons dying of painful tumours who could have been saved by the desert island gold!

"But", the deontologist may still object, "it is my doctrine which is the humane one. You have accused me of inhumanity because I sometimes cause avoidable misery for the sake of keeping a rule. But it is these very rules, which you regard as so cold and inhuman, which safeguard mankind from the most awful atrocities. In the interests of future generations are we to allow millions to die of starvation, or still more millions to be sent to forced labour? Is it not this very consequentialist mentality which is at the root of the
vast injustices which we see in the world today?” Two replies are relevant. In the first place the man who says this sort of thing may or may not be interested in the welfare of future generations. It is perfectly possible not to have the sentiment of generalized benevolence but to be moved by a localized benevolence. When this is localized in space we get the ethics of the tribe or the race: when it is localized in time we get an ethics of the present day and generation. It may well be that atrocities carried out for the sake of a Utopian future repel some people simply because they mortgage the present for the sake of the future. Here we have a difference about ultimate ends, and in this case I cannot accuse my opponent of being either confused or superstitious, though I may accuse him of being limited in his vision. Why should not future generations matter as much as present ones? To deny it is to be temporally parochial. If it is objected that future generations will only probably exist, I reply: would not the objector take into account a probably existing present population on a strange island before using it for bomb tests?

In the second place, however, the opponent of utilitarianism may have a perfectly disinterested benevolence, save for his regard for the observance of rules as such. Future generations may in fact mean as much to him as present ones. To him the utilitarian may reply as follows. If it were known to be true, as a question of fact, that measures which caused misery and death to tens of millions today would result in saving from greater misery and from death hundreds of millions in the future, and if this were the only way in which it could be done, then it would be right to cause these necessary atrocities. The case is surely no different in principle from that of the battalion commander who sacrifices a patrol to save a company. Where the tyrants who cause atrocities for the sake of Utopia are wrong is, surely, on the plain question of fact, and on confusing probabilities
with certainties. After all, one would have to be very sure that future generations would be saved still greater misery before one embarked on such a tyrannical programme. One thing we should now know about the future is that large-scale predictions are impossible. Could Jeremy Bentham or Karl Marx (to take two very different political theorists) have foreseen the atom bomb? Could they have foreseen automation? Can we foresee the technology of the next century? Where the future is so dim a man must be mad who would sacrifice the present in a big way for the sake of it. Moreover even if the future were clear to us, it is very improbable that large scale atrocities could be beneficial. We must not forget the immense side effects: the brutalization of the people who ordered the atrocities and carried them out. We can, in fact, agree with the most violent denouncer of atrocities carried out in the name of Utopia without sacrificing our act-utilitarian principles. Indeed there are the best of act-utilitarian reasons for denouncing atrocities. But it is empirical facts, and empirical facts only, which will lead the utilitarian to say this.

The future, I have remarked, is dim, largely because the potentialities of technological advance are unknown to us. This consideration both increases the attractiveness of a utilitarian ethics (because of the built-in flexibility of such an ethics) and increases the difficulty of applying such an ethics.

Normally the utilitarian is able to assume that the remote effects of his actions tend rapidly to zero, like the ripples on a pond after a stone has been thrown into it. This assumption normally seems quite a plausible one. Suppose that a man is deciding whether to seduce his neighbour’s wife. On utilitarian grounds it seems pretty obvious that such an act would be wrong, for the unhappiness which it is likely to cause in the short term will probably be only too obvious. The man need not consider the possibility that one of his remote descendants, if he seduces the woman, will be a great
benefactor of the human race. Such a possibility is not all
that improbable, considering the very likely vast number
of descendants after a good many generations, but it is no
more probable than the possibility that one of his remote
descendants will do great harm to the human race, or that
one of the descendants from a more legitimate union would
benefit the human race. It seems plausible that the long-term
probable benefits and costs of his alternative actions are
likely to be negligible or to cancel one another out.

An obviously important case in which, if he were a utili-
tarian, a person would have to consider effects into the far
future, perhaps millions of years, would be that of a states-
man who was contemplating engaging in nuclear warfare,
if there were some probability, even a small one, that this
war might end in the destruction of the entire human race.
(Even a war less drastic than this might have important
consequences into the fairly far future, say hundreds of
years.) Similar long term catastrophic consequences must be
envisaged in planning flight to other planets, if there is any
probability, even quite a small one, that these planets possess
viruses or bacteria, to which terrestrial organisms would
have no immunity.

The progress of science and technology could yield many
more cases which might pose dramatic problems to the
moralist. Consider the moral problems which would be set
by a spectacular innovation in the field of positive eugenics,\footnote{Positive eugenics is a matter of encouraging breeding by those with
desirable genes, whereas negative eugenics is a matter of discouraging
breeding of those with undesirable genes. In the present state of knowl-
dge of human genetics, at least, the latter is much more scientifically
respectable than the former. For a spectacular suggestion in the field
of positive eugenics, see the book \textit{Out of the Night} (Gollancz, London,
1936) by the American geneticist H. J. Muller. For a popular account
of the biological difficulties which beset the idea of positive eugenics,
see P. B. Medawar, \textit{The Future of Man} (Methuen, London, 1959),
lectures 3 and 4.}
or perhaps of direct tampering with the human genetic material, or of a spectacular discovery which would enable the life span of man to be prolonged indefinitely. (For example, would the realization of the last possibility imply the rightness of universal euthanasia?) Again, suppose that it became possible to design an ultra-intelligent machine\(^1\) (superior in intelligence to any human) which could then design a yet more intelligent machine which could . . . (and so on).

Consider positive eugenics first. Suppose that it did one day turn out that by methods of positive eugenics, it became possible markedly to increase the intelligence of the whole human race, without using tyrannical or unpleasant means and without reducing the genetic diversity of the species. (There are important biological advantages in diversity.) Ought a utilitarian to approve of such a measure? Clearly something will depend on whether he is a hedonistic or an ideal utilitarian. The ideal utilitarian may have an intrinsic preference for more intelligent states of mind. However the hedonistic utilitarian might agree with the ideal one if he thought that intelligence was extrinsically valuable, for example if he thought that wars and poverty were due mainly to stupidity, and perhaps if he thought that more avenues for obtaining pleasure were open to intelligent people.

Even more interesting ethical issues arise if we imagine that biological engineering went so far as to enable the production of a higher species of man altogether. Similar issues arise also if we imagine that it becomes possible to produce an ultra-intelligent artefact which possesses consciousness. (This is not the place to enter into the deep metaphysical issues which arise out of the question of whether a

\(^1\) See, for example, I. J. Good, ‘Speculations concerning the first ultra-intelligent machine’, *Advances in Computers*, vol. 6, Academic Press, New York, 1965.
10. Utilitarianism and justice

conscious artefact is possible or not.) Let an entity which is either a member of the envisaged superior species or is an ultra-intelligent conscious artefact be conveniently referred to as 'a superman'. What might a utilitarian’s attitude be towards possible actions which would lead to the production of a superman? It is quite possible that there should be a kind of utilitarian who valued only the happiness of his own species and was perfectly indifferent to that of higher and lower species. He might even envisage the superman with fear and hatred. Such a man’s ethics would be analogous to the ethics of the tribe. Suppose alternatively that he were an ideal or quasi-ideal utilitarian, who thought that it was better to be Socrates dissatisfied than a fool satisfied. Should he similarly yield ethical precedence to the superman?

At present there is much less possibility of practical disagreement between those who concern themselves with the happiness of all sentient beings. As regards inferior beings, there is indeed a possibility of serious disagreement over the morality of such things as ‘factory farming’. But if it became possible to control our evolution in such a way as to develop a superior species, then the difference between a species morality and a morality of all sentient beings would become very much more of a live issue.

So far, I have done my best to state utilitarianism in a way which is conceptually clear and to rebut many common objections to it. At the time I wrote the earlier edition of this monograph I did so as a pretty single-minded utilitarian myself. It seemed to me then that since the utilitarian principle expressed the attitude of generalized benevolence, anyone who rejected utilitarianism would have to be hard hearted, i.e. to some extent non-benevolent, or else would have to be the prey of conceptual confusion or an unthinking
adherent of traditional ways of thought, or perhaps be an adherent of some religious system of ethics, which could be undermined by metaphysical criticism. Admittedly utilitarianism does have consequences which are incompatible with the common moral consciousness, but I tended to take the view "so much the worse for the common moral consciousness". That is, I was inclined to reject the common methodology of testing general ethical principles by seeing how they square with our feelings in particular instances.

After all, one may feel somewhat as follows. What is the purpose of morality? (Answering this question is to make a moral judgement. To think that one could answer the question "What is the purpose of morality?" without making a moral judgement would be to condone the naturalistic fallacy, the fallacy of deducing an 'ought' from an 'is'.) Suppose that we say, as it is surely at least tempting to do, that the purpose of morality is to subserve the general happiness. Then it immediately seems to follow that we ought to reject any putative moral rule, or any particular moral feeling, which conflicts with the utilitarian principle. It is undeniable that we do have anti-utilitarian moral feelings in particular cases, but perhaps they should be discounted as far as possible, as due to our moral conditioning in childhood. (The weakness of this line of thought is that approval of the general principle of utilitarianism may be due to moral conditioning too. And even if benevolence were in some way a 'natural', not an 'artificial', attitude, this consideration could at best have persuasive force, without any clear rationale. To argue from the naturalness to the correctness of a moral attitude would be to commit the naturalistic fallacy.) Nevertheless in some moods the general principle of utilitarianism may recommend itself to us so much the more than do particular moral precepts, precisely because it is so general. We may therefore feel inclined to reject an ethical methodology which implies that we should test our
general principles by our reactions in particular cases. Rather, we may come to feel, we should test our reactions in particular cases by reference to the most general principles. The analogy with science is not a good one, since it is not far off the truth to say that observation statements are more firmly based than the theories they test. But why should our more particular moral feelings be more worthy of notice than our more generalized ones? That there should be a disanalogy between ethics and science is quite plausible if we accept a non-cognitivist theory of meta-ethics.

The utilitarian, then, will test his particular feelings by reference to his general principle, and not the general principle by reference to his particular feelings. Now while I have some tendency to take this point of view (and if I had not I would not have been impelled to state and defend utilitarianism as a system of normative ethics) I have also some tendency to feel the opposite, that we should sometimes test our general principles by how we feel about particular applications of them. (I am a bit like G. E. Moore in his reply to C. L. Stevenson, where he feels both that he is right and Stevenson wrong and that he is wrong and Stevenson right. My own indecisiveness may be harder to resolve, since in my case it is a matter of feeling, rather than intellect, which is involved.)

It is not difficult to show that utilitarianism could, in certain exceptional circumstances, have some very horrible consequences. In a very lucid and concise discussion note, H. J. McCloskey has considered such a case. Suppose that the sheriff of a small town can prevent serious riots (in

I say, 'not far off the truth' because observation statements are to some extent theory laden, and if they are laden with a bad theory we may have to reject them.


which hundreds of people will be killed) only by ‘framing’ and executing (as a scapegoat) an innocent man. In actual cases of this sort the utilitarian will usually be able to agree with our normal moral feelings about such matters. He will be able to point out that there would be some possibility of the sheriff’s dishonesty being found out, with consequent weakening of confidence and respect for law and order in the community, the consequences of which would be far worse even than the painful deaths of hundreds of citizens. But as McCloskey is ready to point out, the case can be presented in such a way that these objections do not apply. For example, it can be imagined that the sheriff could have first-rate empirical evidence that he will not be found out. So the objection that the sheriff knows that the man he ‘frames’ will be killed, whereas he has only probable belief that the riot will occur unless he frames the man, is not a sound one. Someone like McCloskey can always strengthen his story to the point that we would just have to admit that if utilitarianism is correct, then the sheriff must frame the innocent man. (McCloskey also has cogently argued that similar objectionable consequences are also implied by rule-utilitarianism. That is, an unjust system of punishment might be more useful than a just one. Hence even if rule-utilitarianism can clearly be distinguished from act-utilitarianism, a utilitarian will not be able to avoid offensive consequences of his theory by retreating from the ‘act’ form to the ‘rule’ form.) Now though a utilitarian might argue that it is empirically unlikely that some such situation as McCloskey envisages would ever occur, McCloskey will point out that it is logically possible that such a situation will arise. If the utilitarian rejects the unjust act (or system) he is clearly giving up his utilitarianism, McCloskey then remarks: “But as far as I know, only J. J. C. Smart among the contemporary utilitarians, is happy to adopt this ‘solution’.” Here I must lodge a mild protest. McCloskey’s use of the word ‘happy’
surely makes me look a most reprehensible person. Even in my most utilitarian moods I am not happy about this consequence of utilitarianism. Nevertheless, however unhappy about it he may be, the utilitarian must admit that he draws the consequence that he might find himself in circumstances where he ought to be unjust. Let us hope that this is a logical possibility and not a factual one. In hoping thus I am not being inconsistent with utilitarianism, since any injustice causes misery and so can be justified only as the lesser of two evils. The fewer the situations in which the utilitarian is forced to choose the lesser of two evils, the better he will be pleased. One must not think of the utilitarian as the sort of person who you would not trust further than you could kick him. As a matter of untutored sociological observation, I should say that in general utilitarians are more than usually trustworthy people, and that the sort of people who might do you down are rarely utilitarians.

It is also true that we should probably dislike and fear a man who could bring himself to do the right utilitarian act in a case of the sort envisaged by McCloskey. Though the man in this case might have done the right utilitarian act, his act would betoken a toughness and lack of squeamishness which would make him a dangerous person. We must remember that people have egoistic tendencies as well as beneficent ones, and should such a person be tempted to act wrongly he could act very wrongly indeed. A utilitarian who remembers the possible moral weakness of men might quite consistently prefer to be the sort of person who would not always be able to bring himself to do the right utilitarian act and to surround himself by people who would be too squeamish to act in a utilitarian manner in such extreme cases.

No, I am not happy to draw the conclusion that McCloskey quite rightly says that the utilitarian must draw. But neither am I happy with the anti-utilitarian conclusion. For
An outline of a system of utilitarian ethics

if a case really did arise in which injustice was the lesser of two evils (in terms of human happiness and misery), then the anti-utilitarian conclusion is a very unpalatable one too, namely that in some circumstances one must choose the greater misery, perhaps the very much greater misery, such as that of hundreds of people suffering painful deaths.

Still, to be consistent, the utilitarian must accept McCloskey’s challenge. Let us hope that the sort of possibility which he envisages will always be no more than a logical possibility and will never become an actuality. At any rate, even though I have suggested that in ethics we should test particular feelings by general attitudes, McCloskey’s example makes me somewhat sympathetic to the opposite point of view. Perhaps indeed it is too much to hope that there is any possible ethical system which will appeal to all sides of our nature and to all our moods. It is perfectly possible to have conflicting attitudes within oneself. It is quite conceivable that there is no possible ethical theory which will be conformable with all our attitudes. If the theory is utilitarian, then the possibility that sometimes it would be right to commit injustice will be felt to be acutely unsatisfactory by someone with a normal civilized upbringing. If on the other hand it is not utilitarian but has deontological elements, then it will have the unsatisfactory implication that sometimes avoidable misery (perhaps very great avoidable misery) ought not to be avoided. It might be thought that some compromise theory, on the lines of Sir David Ross’s, in which there is some ‘balancing up’ between considerations of utility and those of deontology, might provide an acceptable compro-

1 J. W. N. Watkins considers this matter in his ‘Negative utilitarianism’, Aristotelian Society Supp. Vol. 67 (1963) 95-114. It is now apparent to me that my paper ‘The methods of ethics and the methods of science’, Journal of Philosophy 62 (1965) 344-9, on which the present section of this monograph is based, gives a misleading impression of Watkins’s position in this respect.
muse. The trouble with this, however, is that such a ‘balancing’ may not be possible: one can easily feel pulled sometimes one way and sometimes the other. How can one ‘balance’ a serious injustice, on the one hand, and hundreds of painful deaths, on the other hand? Even if we disregard our purely self-interested attitudes, for the sake of interpersonal discussions, so as to treat ourselves neither more nor less favourably than other people, it is still possible that there is no ethical system which would be satisfactory to all men, or even to one man at different times. It is possible that something similar is the case with science, that no scientific theory (known or unknown) is correct. If so, the world is more chaotic than we believe and hope that it is. But even though the world is not chaotic, men’s moral feelings may be. On anthropological grounds it is only too likely that these feelings are to some extent chaotic. Both as children and as adults, we have probably had many different moral conditionings, which can easily be incompatible with one another.

Meanwhile, among possible options, utilitarianism does have its appeal. With its empirical attitude to questions of means and ends it is congenial to the scientific temper and it has flexibility to deal with a changing world. This last consideration is, however, more self-recommendation than justification. For if flexibility is a recommendation, this is because of the utility of flexibility.

ACKNOWLEDGEMENTS

The foregoing is a revised version of my monograph An Outline of a System of Utilitarian Ethics, which was published in 1961 by the Melbourne University Press, with generous financial aid from the University of Adelaide. I should like to reiterate the expression of gratitude, which I made in the preface to the original edition, to the University of Adelaide, and to the officers of the Melbourne University Press, as well as to Professors A. G. N. Flew, R. M. Hare, J. C. Harsanyi, B. H. Medlin, D. H. Monro, and A. K. Stout, who had commented on earlier drafts of the monograph.
An outline of a system of utilitarian ethics

I am grateful to Mrs Patricia Skinner of the Cambridge University Press, for her suggestion that the revised edition be published together with a monograph by Bernard Williams.

In the years since I wrote the original pamphlet I have not been working a great deal in the field of ethics. I have therefore not made very extensive changes in the original. However I have made some attempt to deal with certain criticisms which in the intervening years have been made of utilitarianism in general, and sometimes of my ideas in particular, and I have made various deletions and additions. However, this fails far short of a complete re-thinking of the issues. The final section of this edition is based on parts of my paper 'The methods of science and the methods of ethics', Journal of Philosophy 62 (1965) 344–9.

I should like to acknowledge that over the years since the publication of the first edition I have benefited by suggestions in correspondence or conversation with many friends, including (as well as some of those already mentioned) Professors R. B. Brandt, David K. Lewis, H. J. McCloskey, Richmond H. Thomason, J. W. N. Watkins, Henry West and Bernard Williams, though probably they will feel that I have not benefited enough.

I am also very grateful to Professor R. M. Hare for inviting me to join with him in giving a graduate class on utilitarianism at Oxford in the Michaelmas Term of 1970.

I have added a bibliography, which I hope will be valuable especially for students (including post-graduate students). The literature on utilitarianism has grown to be so vast that I am uneasily aware that I must have omitted some references which ought to be in it, but I have tried to make it at least reasonably detailed.