AN OUTLINE OF A SYSTEM OF UTILITARIAN ETHICS

J. J. C. Smart

HUGHES PROFESSOR OF PHILOSOPHY
UNIVERSITY OF ADELAIDE

MELBOURNE UNIVERSITY PRESS
ON BEHALF OF
THE UNIVERSITY OF ADELAIDE
Smart, John Jamieson Carswell, 1920--

An outline of a system of utilitarian ethics. (Carlton
Melbourne University Press on behalf of the University of

51 p. 22 cm.

Ethics. 2. Utilitarianism. r. Title.

31.S6 171.5 62-4668 ♠
y of Congress [2]
AN OUTLINE OF A SYSTEM OF
UTILITARIAN ETHICS
Preface

I should like to thank several friends who have read and commented on earlier drafts of this monograph, especially Professor A. G. N. Flew, Mr R. M. Hare, Mr J. C. Harsanyi, Mr B. H. Medlin, Professor D. H. Monro, and Professor A. K. Stout. It gives me particular pleasure to express my gratitude to the University of Adelaide for generous financial aid towards the costs of publication, and to the officers of the Melbourne University Press for their co-operation.

J.J.C.S.
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>V</td>
</tr>
<tr>
<td>Introductory</td>
<td>1</td>
</tr>
<tr>
<td>Act Utilitarianism and Rule Utilitarianism</td>
<td>4</td>
</tr>
<tr>
<td>Hedonistic and Non-Hedonistic Utilitarianism</td>
<td>5</td>
</tr>
<tr>
<td>Average Happiness versus Total Happiness</td>
<td>18</td>
</tr>
<tr>
<td>Negative Utilitarianism</td>
<td>19</td>
</tr>
<tr>
<td>Rightness and Wrongness of Actions</td>
<td>20</td>
</tr>
<tr>
<td>The Place of Rules in Act Utilitarianism</td>
<td>29</td>
</tr>
<tr>
<td>Simple Application of Game-Theory Technique</td>
<td>42</td>
</tr>
<tr>
<td>Utilitarianism and the Future</td>
<td>44</td>
</tr>
</tbody>
</table>
INTRODUCTORY

Such writers as J. S. Mill, H. Sidgwick and G. E. Moore produced, as a result of philosophical reflection, systems of normative ethics. Of recent years normative ethics has become distinguished from meta-ethics, the logical analysis of ethical concepts. Indeed, as a result of the prevalence of 'non-cognitivist' theories of meta-ethics, for example those of C. L. Stevenson and R. M. Hare, 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 lie within the field of personal decision, persuasion, advice and propaganda, but not within the field of philosophy as such.

While it is true that some ultimate ethical disagreements may depend simply on differences of ultimate preference (and non-ultimate ethical disagreements depend on differences about empirical facts, which are equally beyond the philosopher's jurisdiction) 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 far too readily assumes that ordinary ethical thinking is relatively clear-headed and worthy of analysis rather than criticism. He thinks of meta-ethics as analogous to the logic of science, whereas it is at least arguable that the conceptual investigation of our ordinary moral notions is more like the philosophy of magic and witchcraft (if there is such a branch of philosophy). At the very least, some practically influential systems of ethical thought depend on metaphysical and theological bases which are a legitimate object of philosophical analysis and criticism.

It will be my object in the present study to state a system of ethics which is free from traditional and theological associations.

---

1 Ethics and Language (Yale University Press, New Haven, 1944).
This is that type of utilitarianism which R. B. Brandt in his recent book calls 'act utilitarianism' and which I, in an earlier article, called, less happily, 'extreme utilitarianism'. 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 will assume for the purposes of this study the truth of some such meta-ethical analysis as that of Hare's Language of Morals. In adopting a non-cognitivist meta-ethics I do, of course, renounce the attempt to prove the act utilitarian system. I shall be concerned more with the clear statement of act utilitarianism in such a way as to show that it can be defended against the frequent, and in my opinion sophistical, objections that are still being brought against it. Nevertheless I should like to indicate my opinion that the choice of conceptually clear and emotionally attractive systems of normative ethics that might be alternatives to it is not as wide as is usually thought.

In the first place, B. H. Medlin has argued that it is impossible to state ethical egoism without either confusion or else a sort of pragmatic inconsistency. If he is right the same sort of thing may quite possibly be the case with other ethical systems besides egoism. Secondly, some widespread ethical systems depend on metaphysical premises, and would be destroyed if their metaphysical bases were cut away from beneath them. I myself would be prepared to argue that this is so in the case of so-called 'natural law' ethics, which depends on a quasi-Aristotelian metaphysics. Thirdly, any system of deontological ethics is open to a persuasive type of objection, which, though it falls short of disproof, can be convincing to those who have the welfare of humanity at heart. For though conceivably in most cases the

---

4 'Extreme and Restricted Utilitarianism', Philosophical Quarterly, vol. 6, 1956, pp. 344-54. In this article I used the terms 'extreme' and 'restricted' instead of 'act' and 'rule' utilitarianism. I now prefer Brandt's terminology.
dictates of a deontological ethics might coincide with those of human welfare and of an act utilitarian ethics, there must be some cases in which they conflict. Indeed there must be some cases in which avoidable human misery could be prevented only by breaking the precepts of the deontological ethics. The adherent of the deontological ethics can then be accused of heartlessness if he prefers abstract conformity with an ethical rule to the prevention of the misery of some fellow creature. Or if not of heartlessness, of some sort of confusion, or perhaps of 'rule worship' which leads him to prefer 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. I should not be particularly concerned to defend utilitarianism against such a deontology, supposing it to exist, for the practical consequences of the deontology would not differ from those of utilitarianism. However all deontological systems known to me differ from utilitarianism not only in theory, but also sometimes in practice.

Such a 'persuasive' objection to deontology is possible simply because we have assumed the truth of non-cognitivist meta-ethics. A cognitivist in meta-ethics like Sir David Ross' could resist this appeal to the heart by saying that whether we like it or not, his deontological principles can be either seen or proved to be true. That they might sometimes conflict with human happiness might seem to him to be of more sentimental than philosophic concern. But if we strip off the cognitivist meta-ethics from Ross's theory his deontology looks artificial and infected with 'rule worship'. It would be harder to produce persuasive arguments, however, against a very restrained deontology, which supplemented the appeal to benevolence by such not obviously 'artificial' sentiments as the desire to repay favours received, to revenge hurts, and to tell the truth, and which made no ultimate appeal to such propensities as to keep promises or repay debts. However I am not attempting the impossible task of 'proving' my system of normative ethics against all rivals: I am merely indicating that it is harder than is com-

monly believed to produce clear-headed and acceptable alternative systems.

In setting up a system of normative ethics I must of course appeal to some ultimate attitudes which I must suppose to be held in common by me and by those to whom I am addressing myself. The sentiment to which I will appeal will be generalized benevolence: the disposition to seek the happiness of mankind (or perhaps better, of all sentient beings) generally, not counting one’s own happiness more or less than that of any other man. This is a sentiment which arises from sympathy: I shall be assuming that my audience consists of benevolent and sympathetic men. Of course such an audience will have other propensities besides that of benevolence. For example, they may have a propensity to obey the rules of some traditional moral code irrespective of whether such an action is that which would be dictated solely by generalized benevolence. Nevertheless I may have some good hope of getting such an audience to agree with my system of normative ethics after I have made some philosophical clarifications. They may come to suspect that their non-utilitarian propensities depend on conceptual confusion. In making its ultimate appeal to such a natural and widespread human sentiment as that of generalized benevolence, act utilitarianism is in a very strong position. That it will not be accepted by everybody is of course no more an objection to it than to any other ethical, or even scientific, doctrine.

**ACT UTILITARIANISM AND RULE UTILITARIANISM**

The system of normative ethics which I shall be concerned to defend is, as I have said earlier, *act* utilitarianism, not *rule* utilitarianism. Act utilitarianism states that the rightness or wrongness of an action is to be judged by the consequences, good and bad, of the action itself. Rule utilitarianism holds 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. Toulmin, and with the
latter, one like Kant's. 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'.

I have argued elsewhere the objections to rule utilitarianism as compared with act utilitarianism. Briefly they boil down to the accusation of rule worship: 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 'everyone 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.

**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?

---

tham,\textsuperscript{14} 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,\textsuperscript{15} 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.\textsuperscript{16} 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’ to be a state of affairs brought about by keeping 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 ethics that I am assuming, the function of the words ‘ought’ and ‘good’ is primarily 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


\textsuperscript{16} J. S. Mill, \textit{Utilitarianism}, Everyman ed.
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 our 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 people I am trying to convert will mostly agree with me about what consequences are good, but will disagree with me about whether we ought always to do that which will produce the best consequences. For a reason, which will appear presently, the differences between ideal and hedonistic utilitarianism in most cases will not 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'. But 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 catalyst 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 invent something or discover 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 sunbathing 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 fecund 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, 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. And with the increase of modern science we need ever increasing thought and labour to survive. To take only one example, as cures are found for diseases more unfit people survive, more people of poor heredity are propagated, more medical science is required, and so on. We
need all the intelligence we can muster to keep one jump ahead of the consequences of our own intelligence. No wonder that many of those who survive 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 the allegedly moronic fodder of commercial television. But even the most avid television addict probably enjoys solving practical problems connected with his car, his furniture, or his garden. However un-intellectual 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. 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 science, practical arts, and the physical and mental well-being 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. Recently 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 characteristic 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 products 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 certainly so with the ordinary sensual pleasures. Excessive indulgence in the physical pleasures of sex does have a debilitating effect and does 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, 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, for 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
of Utilitarian Ethics

a certain state from tomorrow onwards, even though we may know that from tomorrow onwards we shall be perfectly satisfied. All this is psychologically possible. It is just the obverse of a situation that we often find. Thus yesterday, as I first wrote this, I was suspended by cable car half-way up a precipitous mountain. While I was in the cable car and as I looked at the yawning chasm below I fervently wished that I had never come. 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, 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 shall 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 a situation. 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 is 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 philosophic 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 contentment 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, 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 listen-

ing to a lecture on philosophy. Even sensual pleasures come under the same description. For example the pleasure of eating an ice-cream essentially involves having a certain physical sensation, in a way in which the pleasures of golf or symbolic logic do 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 ball as much as 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 operators 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 in the final section of this study. 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 differences 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
An Outline of a System

the less complex ones: they are not only enjoyable in themselves but 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 replacing 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 had been given an electric shock whenever we had tasted cheese, then cheese would have been immediately distasteful to us. Our repugnance to the sadist arises, naturally enough, be-

cause in our universe sadists invariably do harm. If we lived in a universe in which by some extraordinary laws of psychology a sadist was always confounded by his own knavish tricks and invariably did a great deal of good, then we should feel better disposed towards the sadistic mentality. Even if we could de-condition ourselves from feeling an immediate repugnance to a sadist (as we could de-condition ourselves from a repugnance to cheese by going through a course in which the taste of cheese was invariably associated with a pleasurable stimulus) language might make it difficult for us to distinguish an extrinsic distaste for sadism, founded on our distaste for the consequences of sadism, from an immediate distaste for sadism as such. Normally when we call a thing ‘bad’ we mean indifferently to express a dislike for it in itself or to express a dislike for what it leads to. When a state of mind is sometimes extrinsically good and sometimes extrinsically bad, we find it easy to distinguish between our intrinsic and extrinsic preferences for instances of it, but when a state of mind is always, or almost always, extrinsically bad, it is easy for us to confuse an extrinsic distaste for it with an intrinsic one. If we allow for this, it does not seem so absurd to hold that there are no pleasures which are intrinsically bad. Pleasures are bad only because they cause harm to the person who has them or to other people. But if anyone likes to disagree with me about this I do not feel very moved to argue the point. Such a disagreement about ultimate ends is not likely to lead to any disagreement in practice. For in all actual cases there are sufficient extrinsic reasons for abhorring sadism and similar states of mind. Approximate agreement about ultimate ends is often quite enough for rational and co-operative moral discourse. In practical cases the possibility of factual disagreement about what causes produce what effects is likely to be overwhelmingly more important than disagreement in ultimate ends between hedonistic and ideal utilitarians.

There are of course many valuations other than that of the intrinsic goodness of sadistic pleasures which divide the ideal from the hedonistic utilitarian. For example the ideal utilitarian would hold that an intellectual experience, even though not pleasurable, would be intrinsically good. Once more, however, I think we can convince ourselves that in most cases this disagree-
ment about ends will not lead to disagreement about means. Intellectual experiences are in the hedonistic view extrinsically good. Of course there may be wider issues dividing the hedonistic from the ideal utilitarian, if Moore is the ideal utilitarian. I would argue that Moore's principle of organic unities destroys the essential utilitarianism of his doctrine. He need never disagree in practice, as a utilitarian ought to, with Sir David Ross. Every trick Ross can play with his prima facie duties, Moore can play, in a different way, with his organic unities.

**AVERAGE HAPPINESS VERSUS TOTAL HAPPINESS**

Another type of ultimate disagreement between utilitarians, whether hedonistic or ideal, arises over whether we should try to maximize the *average* happiness of human beings (or the average goodness of their states of mind) or whether we should try to maximize the *total* happiness or goodness. (I owe this point to Professor A. G. N. Flew.) I have not yet elucidated the concept of total happiness, and you may regard it as a suspect notion. But for present purposes I shall put it in this way: would you be quite indifferent between (a) a universe containing only one million happy sentient beings, all equally happy, and (b) a 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. 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.
K. R. Popper has suggested\(^{21}\) 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 we found a new university. We may hope that indirectly research will help to minimize pains, but that is not the only 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.\(^{22}\) 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 easy to show 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.\(^{23}\) 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.

Though I am not attracted to negative utilitarianism as an ultimate principle, I do concede that the injunction 'worry about

\(^{23}\) Popper, op. cit.
removing misery rather than about promoting happiness' has a good deal to recommend 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.

RIGHTNESS AND WRONGNESS OF ACTIONS

I shall now state the act utilitarian doctrine. I shall put it forward in a broadly hedonistic form. If anyone values states of mind such as knowledge independently of their pleasureableness 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. This is so simple and natural a doctrine that we can surely expect that my readers, or at any rate those readers with whom I am interested in talking ethics, 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 my hypothesis, then, that 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, then there will be little need to defend the utilitarian 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, I cannot appeal to argument and must rest my hopes on the good feeling of my readers. If any reader is not a sympathetic and benevolent man, then of course I cannot expect him to have an ultimate pro-attitude to human happiness in general.

My 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 in that it would 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 our 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
An Outline of a System

order to help him decide whether to do A or B we could say to a man: '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 would be absurd to deny 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 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 you had some reason for knowing that every genera-
tion would be more happy than not you would not need to be
worried that the remote consequences of your actions would be
in detail unknown. The necessity for the 'ripples in the pond'
postulate comes from the fact that 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 descendants 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 probabilities
An Outline of a System

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 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 high 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 high-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 of opinion 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 inequalitarian parent would get the other to agree with him if he could convince him that there was a much higher probability of Etonians 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 important sort of disagreement, to have intelligent moral discussion with a person we do not in fact need complete agreement with him about ultimate ends: an approximate agreement is sufficient.

Rawls 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. As against this suggestion I would make the

---


25 See especially p. 168 of Rawls's article.
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 I concede that my 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 utilitarianism that it involves the allegedly absurd notion of a 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 quantitative, judgment 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 probability of a slightly good result the year after, while another action may give a very small probability of a superlatively good result next year with a quite 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 con-
ceptual difficulty, but it now appears that we have to go within total situations and weight different elements within them according to different probabilities. 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 total situation method. 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 toothache 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 $2/5$ 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 $1/5$. 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 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 attitudes, 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 advantage or a probable small advantage, and in these cases perhaps we shall have to toss a penny in deciding 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 authoritarianism, dogmatism or romanticism.

So, at any rate, it appears at first sight. But if I cannot say any more my position has a serious weakness. My method of developing normative ethics is to appeal to feelings, namely of benevolence, 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 approximate 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 sub-
jective 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 unbiassed and far-sighted man. This, however, would 
require independent criteria for lack of bias and for far-sighted-
ness. 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 situ-
ation may not be hopeless. But until we have an adequate theory 
of objective probability utilitarianism is not on a secure theoreti-
cal 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 deonto-
logical ethics implies some method of weighing up the claims of 
conflicting prima facie duties, for it is impossible that deonto-
logical rules of conduct should never conflict, and the rationale 
of this is even more insecure than is the theory of objective 
probability.

THE PLACE OF RULES IN ACT UTILITARIANISM

I hold, then, that 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

26 Decision Making (Stanford University Press, Stanford, California, 1957).
28 R. McNaughton's interesting article 'A Metrical Concept of Happiness', Philosophy and Phenomenological Research, vol. 14, 1953-4, pp. 171-83, does not enable us to propose a complete utilitarian calculus, because it neglects probability considerations.
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. 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 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
of Utilitarian Ethics

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. 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 my utilitarian doctrine incompatible, as M. A. Kaplan 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 where we do indeed decide what to do. On those 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

29 Morton A. Kaplan, "Some Problems of the Extreme Utilitarian Position", Ethics, vol. 70, 1960, pp. 228-32. This is a critique of my earlier article. He also puts forward a game theoretic argument against me, but this seems cogent only against an egoistic utilitarian.
or non-utilitarian, comes into the matter. What 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, 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 the right action: there are two or more actions which are 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 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 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 as 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 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. Our 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 our 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.

I shall now make a terminological recommendation. I propose to 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, I suggest that we 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 the highly probable though not quite so good results.
a more accurate formulation we should have to weight the goodness of the results with their probability. 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 of 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. In effect, it is rational to do what you reasonably think to be right, and what will be right is what will have the best results. 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. 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, a bad motive is one which generally ends in maleficent actions. Clearly there is no incon-
sistency 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 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.

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 did 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, was 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 shall we praise the action (thus expressing our approval of it) but we shall 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
An Outline of a System

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 optimific. 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 he 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 we are 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 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 situa-
tion. 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: 31

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, 'any particular existing moral rule, though not the ideally best even for such beings, as existing men under the existing circumstances, may yet be the best that they can be got to obey'. 32 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 again, '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'. 33

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 I have recommended. When a man says that another is wicked he usually is saying something of a partly metaphysical or superstitious connotation. He is saying that there is something like a yellow stain on the other man's soul. Of course he does 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 are left in comfortable obscurity. Nevertheless the things he does say do 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 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. Prior\textsuperscript{34} quotes the nursery rhyme:

\begin{verbatim}
For want of a nail
The shoe was lost;
For want of a shoe
\end{verbatim}

\textsuperscript{34} Aristotelian Society Supplementary Volume 30, 1956, p. 95.
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 will make him neurotic and even more horses in future will 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 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 prevented 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, 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 a criticism of act utilitarianism should criticize it as a normative system. Some writers seem to think that to refute utilitarianism it is sufficient to show

that it conflicts with some of our particular moral intuitions or feelings. R. B. Brandt and H. J. McCloskey have both taken this line. A false analogy with science seems to operate. In science we certainly must correct our general principles in the light of particular observations. But in ethics our particular statements are not observation reports but recommendations. It therefore seems to me that the matter is the other way round. Our general principle, resting on something so simple and natural as generalized benevolence, seems to me to be more securely founded than our particular feelings, which are subtly distorted by analogies with similar looking (but in reality totally different) types of cases, and with all sorts of hangovers from traditional and uncritical ethical thinking.

If, of course, act utilitarianism were put forward as a descriptive systematization of how the ordinary man, or even we ourselves, in our unreflective and uncritical moments, actually thinks 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 the sort of doctrine which I shall be advocating is destroyed and no part of it left standing. 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, by implication, with what they ought to be. The correctness 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.

SIMPLE APPLICATION OF GAME-THEORY TECHNIQUE

So far I hope 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 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 considers the case of a utilitarian in wartime England, and it is supposed that there is a governmental request that a maximum temperature of $50^\circ$ 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 $70^\circ$ 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.

The act utilitarian will have to agree that if his behaviour could be kept secret then he ought in this case to use the electricity and gas. But he 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. The utilitarian calculation would indeed be different if we assumed that the behaviour of the members of the Cabinet would leak out. This objection to utilitarianism depends on a failure to keep distinct the question of the utility of an action and utility (especially in a non-utilitarian society) of praise or blame of it.

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 premiss in his calculation the proposition that very

---

few people would be likely to reason as he does. They would presumably be adherents of a traditional, non-utilitarian morality.

How would the Frenchman reason if he were living in a society composed entirely of convinced 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 they know 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 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 I have in mind \( p \) will be so near 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 \( p \) could be obtained. This type of reasoning seems to me important more for the theoretical insight it affords than for its potentiality for practical guidance.

It may be thought that my symmetrical solution by means of mixed strategies implies that my utilitarianism is after all not act but rule utilitarianism, of the sort that I have called Kantianism. This is not so for two reasons. In the first place, the symmetrical adoption of a mixed strategy, is, if it is Kantianism, a very different sort from the usual 'all or none' variety. In the second place, even such a modified form of Kantianism would still imply conduct different from that of the act utilitarian in those cases in which the agent is living in a largely non-utilitarian society.

**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

\[40\] 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'. I should have used it myself instead of the thoroughly confused reasoning at the top of p. 352 of my earlier article.
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 (remember the promise was known only to me, so my action will not in this case weaken the general confidence in the social institution of promising) without being open to the charge of heartlessness? 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 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 I 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 a programme of liquidation of peasants, or whatever else we thought would help to bring about Utopia. And 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 them. But it is empirical facts, and empirical facts only, which lead me to say this.

The future, I have said, is dim because the potentialities of technological advance are unknown to us. This both increases the need for a utilitarian ethics and increases the difficulty of applying such an ethics. Science seems to be progressing with an exponentially increasing rapidity. We are already aware of a rapid change in our physical environment, and no doubt it will not be long before applications of human biology will have even more striking effects. They will transform, to some extent, the very nature of man himself. Less and less will a traditional
and dogmatic ethics be relevant to our needs. There will be a continual necessity for a thoroughgoing re-investigation of ethical issues.

Such a task is not likely to be an easy one. 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 a man is deciding whether to seduce his neighbour's wife. On utilitarian grounds he should quickly come to the conclusion that such an act would be wrong, for the unhappiness which it is likely to cause is only too obvious. He 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 a very improbable one, and in any case it is no more likely than that a remote descendant may do great harm to the human race. The long-term consequences of his actions are likely to be negligible or to cancel one another out.

Sometimes, however, we do not seem to be able to disregard the long-term consequences of our actions. Consider the question of medical and social services. Are these a good thing? At first sight the answer seems only too obvious that they are. But look again. Are we not vastly increasing the proportion of people among our descendants who suffer from hereditary diseases and weaknesses? Are we not also increasing the proportion of people among our descendants who are lacking in resourcefulness and self-reliance? In the bad old days a man with a weak heart and with criminal tendencies and a dislike of work would not live very long and would not propagate many children. Now he is cared for in state hospitals, kept at state expense in prison for a time and then let out and either found a job or maintained on the dole. All this is very humanitarian, but is not its long-term effect disastrous? My own opinion is that the long-term effect is not disastrous: that medical science will be able to cope with the bad physique of the man's descendants. After all, the progress of medicine may be given by an exponential curve. In any case, before the human race could degenerate very markedly we will

doubtless have increased our control over human heredity to such an extent that the degeneration will be more than offset. The progress of science is shown by an exponential curve, and maybe it is still safe to go on extrapolating. As against this, someone might argue that it is not safe to go on extrapolating: that the inevitable increase in specialization will cause different branches of science to lose contact with one another, and an excess of information will cause the channels of communication to get blocked. In any case, it may be said, the liability to hereditary weaknesses may itself be increasing exponentially. I do not therefore know how to answer the question of whether medical and social services are really as humane as they look: I think that probably they are, but they may well not be. That a utilitarian may have to confess doubt and ignorance is of course in accordance with his empirical attitude: the traditionalist or dogmatist in ethics may think that he knows the answer to all questions, but it is salutary to become aware of our limitations.

Eugenics is another field of human biology which raises interesting ethical problems. Let us suppose that it becomes possible to apply techniques of positive eugenics for increasing the intelligence, health and physique of the human race, or a part thereof. Probably as yet we know all too little of human genetics and also of the possible sociological effects of eugenic measures, but let us nevertheless consider the possibility of spectacular measures such as those which were suggested in 1936 by the distinguished American geneticist H. J. Muller. Let us concentrate on the possibilities of increasing average human intelligence.

Now there is a good deal to be said for the view that if the human race were very much more intelligent it would ipso facto be very much more happy. Most people's interests would be in the sciences and the arts which transcend national boundaries, the catch-cries of demagogues and of political parties as we know them would seem utterly childish and barbaric, automation would give abundant leisure which would be spent in intellectual conversation and physical exercise, and the millennium would have arrived. But perhaps it would not have arrived. Perhaps the very intelligent people would be no happier than we

are, in the sense that they would be just as discontented and would enjoy themselves no more. After all, we are temperamentally suited to our own barbaric environment. Let us, then, consider the hypothesis that to increase human intelligence in the manner suggested would not be to increase human contentment and enjoyment. Would this show that we ought not to increase human intelligence? Two utilitarians might well find themselves in ultimate disagreement on this issue. This would happen if they were in ultimate disagreement over the question of Socrates dissatisfied and the fool satisfied. But they would come to agreement if it could be shown that the eugenic millennium would be desirable on purely Benthamite grounds.

Looking even farther into the dim and distant future let us suppose that, centuries hence, biological science enables us to induce favourable mutations in such a way as to enable us to develop a species which is superior to ours, much as ours is to the ape. Would this be something for which we should strive? Further possibilities of ultimate disagreement turn up here. It is quite possible that a man should not only be indifferent to the happiness of both superior and inferior species, but positively hostile to other species. He might envisage the superman with fear and hatred. His ethics would then be, on a grand scale, analogous to the ethics of the tribe. At present there is no possibility of large-scale practical disagreement between those who concern themselves with the happiness of the human species and those who concern themselves with the happiness of all sentient beings. As regards inferior animals, practical issues are of slight importance. Maybe there would be a difference of opinion about the morality of fox-hunting. And in South Australia it is frequently the practice to shear sheep in the winter, so that many die of cold and more are acutely miserable. No doubt one who was concerned with the happiness of humanity only would differ on these questions from one who was concerned with all sentient beings. On the whole this sort of disagreement would not matter, because usually the best way to increase the happiness of all sentient beings is to work for the happiness of human beings. But if it became possible to control our evolution in such a way as to develop a superior species, then the differences between a species morality and a morality of all sentient beings
would become important. For the question of whether we are concerned with the good of the human species or whether we are concerned with wider aims will arise. This question also has a certain persuasive force even if it is put up as a merely theoretical possibility. If a person is inclined to rate the fool so much lower than Socrates, would he also be inclined, as consistency dictates, to rate his own happiness as negligible compared with that of some superior species?

This is the important point that lies behind the so-called ‘evolutionary ethics’ which was so popular towards the end of the last century and which is represented today by men like Sir Julian Huxley and C. H. Waddington. Their writings have tended to be neglected by professional philosophers because of technical defects in logic. They make the mistake of supposing that if the direction of evolution is from A to B then B must be better than A, whereas to say that B is better than A is to express a more favourable attitude to B than to A. But we could equally have expressed a more favourable attitude to A than to B. It just depends on what we like. Nevertheless, as I have said, it is an empirical fact that many of us like complexity. Moreover those who like complexity will inevitably come to dominate those who do not. If it were in our power to develop a superior intelligence, some of us might desire to develop it and some might desire not to do so. It would not matter. Those who did develop it would inevitably enslave those who did not. If ever we are able to make the choice between developing the superman and not developing him, the choice will not be of great importance to any individual. Sooner or later someone will choose in favour of evolutionary advance and this choice will be irreversible.

The production of a higher species is a possibility for the so far distant future that it is perhaps fanciful to discuss it at all. Nevertheless discussion of it helps to bring out the possibilities of ultimate ethical disagreement among different types of utilitarians, and also the very real re-thinking of ethics that the biological sciences might force on us. In a lesser degree this is

sure to happen in the not so distant future from less spectacular, but still startling, possibilities in the field of human biology.

If such questions overwhelm us with their immensity, it is perhaps comforting to reflect, as G. Wald has done, that according to certain plausible cosmological theories there may well be millions of millions of planets similar to our own scattered in inaccessible depths of space. If life on earth should fail, the struggle will still go on in other places. All this is, of course, not much more than guesswork. But if it is true, then on a long-term and large-scale view our human efforts do not much matter. Nevertheless on a small-scale view they matter very much indeed. If I did not believe this I should not have written this monograph.