FORMS AND LIMITS OF UTILITARIANISM

DAVID LYONS

OXFORD
AT THE CLARENDON PRESS

FORMS AND LIMITS OF UTILITARIANISM

This book has been printed digitally and produced in a standard specification in order to ensure its continuing availability

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Singapore Taipei Tokyo Toronto

with an associated company in Berlin

Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

Reprinted 2002

ISBN 0-19-824197-6

TO THE MEMORY OF

LORRAINE HANSBERRY

PREFACE

TELEOLOGISTS claim that the rightness of acts depends solely upon their utility, that is, upon their contribution towards intrinsically good states of affairs. For example, the Act-Utilitarian holds that an act is right if, and only if, its effects are (or are likely to be) no worse than the effects of the alternatives open to the agent. Deontologists deny this; they maintain that the rightness of acts is not simply a function of their utility. They contend that acts are right or wrong because they are acts of this or that kind. They argue, typically, that right acts, regardless of their good or bad effects, must conform to moral rules. And it has often seemed that these two types of moral theory are incompatible because acts of a given kind can vary widely in respect of their utilities; or because acting in accordance with generally acknowledged moral rules sometimes has worse effects than breaking them.

Thus there is a recurring conflict in ethical theory between the partisans of *Utility* and the supporters of *Obedience to Rules*. Attempts have of course been made to reconcile these two factions. The most recent efforts have issued forth a child of both houses: rule-utilitarianism. We may find foreshadowings of rule-utilitarianism in the classical utilitarians, Hume, Bentham, Mill, and Sidgwick. But only within the past two or three decades has the child come of age, clearly formulated as an alternative, proposed and dedicated to a reconciliation, self-consciously unique and distinct from the traditional types of teleology and deontology.

As its name suggests, rule-utilitarianism assigns places to both utility and rules: generally speaking, acts are to be regarded as right only if they conform to rules that can be supported on utilitarian grounds. The utility of an individual *act* is not considered; but a *rule* requires utilitarian justification. Proponents of rule-utilitarianism have argued that the defects of both types of traditional theory are avoided while their virtues are preserved. Thus rule-utilitarianism has received favourable notices as a promising ethical theory, and to some it has seemed the most

plausible compromise between teleological and deontological theories of right conduct.

It once seemed so to me. I thought rule-utilitarianism could profitably be applied in the criticism of social rules, laws, and institutions. But two features of the literature on rule-utilitarianism gave me pause. First, there is general disagreement over the details of rule-utilitarianism. Each theorist has his own version —which, perhaps, is only to be expected. However, some theories that have been considered 'rule-utilitarian' make no more than incidental reference to rules and lay stress instead upon a utilitarian form of generalization test, 'What would happen if everyone did the same?' Other theories emphasize an appeal to rules as such. The differences between these two types of 'rule-utilitarianism' have hardly been noted in the literature. There are, moreover, considerable variations upon both themes. Among the theories stressing rules, for example, some are less purely 'utilitarian' than others; some are formulated in terms of 'ideal', maximally useful rules while others place a premium upon socially accepted, conventional rules and standards of conduct.

The consequent number and variety of actual and possible versions of 'rule-utilitarianism' prove an embarrassment to one who simply wishes to apply the theory. How shall one proceed? Should one select among the versions already proposed, or formulate a new theory for oneself? I found myself forced to step back and take a fresh look, to see what features the various forms of rule-utilitarianism had in common, and to inquire whether rule-utilitarianism could possibly accomplish all that its proponents had hoped.

The second disconcerting feature of the rule-utilitarian literature can be suggested by means of an example. Suppose the utilitarian wishes to determine whether it would be wrong not to vote when voting entails some minimal sacrifice or hardship. Now our ordinary notion of political obligation suggests that it would be wrong not to vote, except in special circumstances. But having to accept some minimal sacrifice or hardship does not seem to be one of those special circumstances. It is supposed that, since the hardship can be avoided by not voting, with little or no effect on the outcome, the Act-Utilitarian would condone

not voting in such a case. But if we apply the generalization test, we find that it would be wrong not to vote because of the bad effects of *everyone's* doing the same. So far so good—for this suggests that a revised utilitarianism is strengthened in a way some philosophers desire, that it is more in accordance with our ordinary notions of particular obligations.

But many rule-utilitarians qualify the judgement that it would be wrong in such circumstances not to vote. 'If', some say, 'it is likely (or one has reason to believe, or one knows) that others will not vote, and therefore that the bad effects which could be avoided by everyone's voting will occur anyway, then it would not be wrong to abstain in order to avoid the hardship.' In allowing such a qualification, however, the rule-utilitarian seems to surrender the substantive difference between the new and the old forms of utilitarianism. Moreover, the argument for the qualification seems inconsistent with the purely suppositional character of the test originally employed. First we asked, 'What *would* happen *if* everyone failed to vote?' Now we ask whether everyone will *in fact* fail to vote.

The first difficulty mentioned (the diversity of theories) appears the more basic. I therefore sought to discover what rule-utilitarianism *is*—not in one arbitrary form, but in its central features. I sought a basis for analysing this body of theories, some common denominator. (In the course of the analysis I found also that a solution to the second problem—whether and in what respect others' actual behaviour should be considered—was crucial to an understanding of rule-utilitarianism.)

I found it helpful, first, to concentrate upon the notion of generalization. For one can readily contrast the simple (or more traditional) utilitarian test, 'What will happen if *this* act is performed?', with the generalization test, 'What would happen if *everyone* did the same?' One can specify two classes of principles, simple and general utilitarian principles, the members of which can be paired as strictly analogous, differing only in the relevant respect. (Among the former we find Act-Utilitarianism, the heir-apparent of traditional utilitarianism as that is currently understood. Thus we can contrast the analogous general utilitarian principle in order to see whether it too is—or can be—subject to the criticisms offered against Act-Utilitarianism.)

Moreover, from the notion of generalization one can begin to build up to the notion of a rule grounded in utility, thus incorporating into the analysis rule-utilitarianism properly so-called.

This book is the outcome of the analysis. I have tried to examine several *types* of moral theory and also to preserve contact with the relevant literature, that is, with theories that have been offered and discussed by philosophers. But these objectives are sometimes difficult to reconcile, since not all current theories match exactly the pattern of principles I have chosen to emphasize. I have always decided in favour of the first objective: the present work is a systematic study of moral theories of certain determinate descriptions rather than a survey and criticism of the contributions of a number of theorists.

For the sake of those who prefer to know in advance where I shall lead them (and perhaps whether the trip is worth the effort), I offer this summary of my main conclusions. I shall argue (1) that there is no essential problem about the generalization test itself, for example, how to determine which acts are 'the same'; an analysis of utilitarian generalization provides the criteria for determining relevant similarities and dissimilarities among acts when the generalization test is given a purely utilitarian interpretation; (2) that, contrary to widespread misapprehensions, two formally different kinds of utilitarianism, simple and general, and along with the latter one kind of rule-utilitarianism, are extensionally equivalent; that is, analogous principles of the various kinds necessarily yield equivalent judgements in all cases; or, in other words, it makes no difference in theory whether the simple or generalization test is applied to acts or—within limits—whether an appeal is made to rules grounded in utility; (3) that other rule-utilitarian theories, those *not* extensionally equivalent to simple utilitarianism, are, ironically, different from it in the wrong ways; (4) that any appeal to generalization or to rules consequently fails to escape the force of traditional arguments against utilitarianism; in particular (4*a*) that in some contexts the supposed force of a generalization test can be accounted for only by strictly non-utilitarian arguments from justice or fairness; and (4*b*) that other 'prima facie duties' recognized by deontologists, such as fidelity, cannot plausibly be accounted for by reference to utility alone.

The point of the title is now clear. We are dealing with paradigms of various forms of utilitarianism. We find that little is gained by choosing one form rather than another. We find also that no pure utilitarian theory can account for some of our strongest moral convictions.

The thesis of extensional equivalence, that is (2), seems to me the main result. This is because the contrary thesis has generally been assumed—an assumption shared by this writer when he embarked upon the present study. The development of this thesis is worth remarking upon in one respect. I have tried to cut through the mode of argument commonly employed in comparing alternative theories of right conduct, argument by example and counter-example. I have tried to press the forms of utilitarianism to their logical conclusions in order to extract their implications schematically. Consequently, my argument purports to be conclusive. If it is correct, the thesis is necessarily true. Its import is restricted only by my analyses of the principles in question, and these the reader may judge for himself.

It may be helpful to mention, in order to forestall misunderstanding, that a few concepts of general theoretical importance are discussed at length from a certain point of view. The concept of an action and its morally relevant description, the notion of a moral rule, and the idea of *ceteris paribus* principles are cases in point. Now it is commonly held that 'meta-ethical' questions are logically prior to questions in 'normative' ethics, that answers to the former are presupposed by answers to the latter. Thus one might expect an analysis of such concepts as I have mentioned logically to precede and to be *independent* of an analysis of the forms of utilitarianism. My argument does not reflect any such presumption of logical priority. I am not convinced that the notion of logical priority, so understood, is appropriate here. If I were propounding my own moral theory I should give precedence to an analysis of 'action' and 'rule'; but I am not here propounding my own moral theory. I am trying to bring to the surface implications of utilitarianism. Thus I have attempted to explicate the use of these concepts within the framework of utilitarianism. And I would suppose that one test of the adequacy of a moral theory, such as utilitarianism, is how such concepts are handled—or mishandled—within it.

This book grew in two stages, as a doctoral dissertation written while I was at Harvard University and presented there in 1963, which was considerably revised and expanded during a year in Oxford. I incurred sizeable debts of gratitude to teachers and friends at both places for criticism and encouragement, debts I am happy to acknowledge here: to H. L. A. Hart, John Rawls, Roderick Firth, William Frankena, David Hodgson, Henry West, and David Kurtzman. Many of their ideas and suggestions have been incorporated into my argument. I have learned from them more than my words may suggest and also from the writers mentioned and discussed in the following pages. But in all these cases, only I can be blamed for the use to which these ideas have been put and for the total consequences.

During the periods of writing, a stable economy was provided by Harvard University through fellowship grants, including a Frank Knox Memorial Fellowship, which made the year in Oxford possible; and by the Woodrow Wilson National Fellowship Foundation through a Dissertation Fellowship. A Faculty Research Grant from Cornell University helped in preparing the book for publication.

Karen Johnson contributed her time and skill to provide a typescript of the final draft. Richmond Campbell and K. Codell Carter assisted in proof-reading and indexing. My son Matthew provided an atonal musical background during the second stage of writing. And my wife Sandra does not cease to amaze me in her ability to tolerate and encourage a philosophizing spouse.

D. L.

CONTENTS

I

UTILITARIAN GENERALIZATION

SOMETIMES an act is criticized just because the results of everyone's acting similarly would be bad. The *generalization test*, 'What would happen if everyone did the same?' is often used in raising such criticisms; and a principle warranting the criticism is of the following kind:

(G1) If the consequences of everyone's doing a certain sort of thing would be undesirable, then it would be wrong for anyone to do such a thing.

This principle is clearly teleological (utilitarian) since in appealing to it, in determining whether acts are wrong, we consider only desirable and undesirable effects—their utility. It is also a generalization principle: the consequences of a general practice (everyone's doing the same) are considered; a particular act is assessed as an act of that kind; and thus the verdict applies to all such acts. Such a principle may therefore be called a form of *utilitarian generalization*.

Challenges employing the generalization test are not uncommon in everyday moral argument. The significance of the test has, however, puzzled philosophers. Thus the subject is of some interest in its own right. But the generalization test has philosophical importance today primarily because it has been associated with the forms of utilitarian generalization, and because these principles, in turn, have seemed to accomplish certain moral tasks on strictly utilitarian grounds which other forms of utilitarianism fail to do.

Our main subject then is this family of principles—their logic and their substantive import: the members of the family; how they compare with the more traditional kind of utilitarian principle; how they may properly be applied. Moreover, I shall use this inquiry—the methods of analysis which I shall adopt and the conclusions that are reached—as a basis for examining a much-heralded recent theory in this tradition, rule-utilitarianism.

I shall relate all these principles to considerations of justice and fairness—to show that the apparent force of the generalization test requires appeal to more than utility.

In this first chapter I shall identify and characterize utilitarian generalization, suggest difficulties and issues, sketch an historical framework, and indicate the dimensions within this class of principles.

A. *Two Kinds of Utilitarianism*

'Oh look!' she said, pointing off to the right. 'The apples are ripe in that orchard. Let's stop and pick some.'

'No. . . .' He drove on, more slowly. 'I don't think we should. Suppose everyone did that!'

'Don't be silly—not everyone will. And the few we'd take wouldn't be missed.'

'But that's beside the point. If we can do it then so can anyone else. And if everyone did the same . . .'

And if everyone did the same, if every passer-by picked as he chose, this grower (or perhaps all growers) would suffer irretrievable losses. Moreover, he might ask himself: 'Does it pay to take such care of my orchards if others are to pick them bare?' Thus, his incentive could be undermined and future production could thereby be damaged. Or he might be obliged, at considerable cost, to post guards and erect fences that would mar the now pleasant landscape.

If such contingencies were the ground for our moralizer's objection, then he was employing the generalization test. He was appealing to a form of utilitarian generalization, such as (G1).

Notice how our moralizer did not argue. He did not claim that the grower would suffer hardship or loss as a result of the small expropriation proposed by his companion. Nor did he say that such hardship or loss would indirectly flow from the act, as a result of their example inciting others to do likewise, sparking a chain reaction leading to a devastation of the orchard. Nor did he maintain that in doing such a thing he and his companion were disposing themselves to act in future in ways which ultimately would have bad consequences. Finally, our moralizer did not mention the contingency, the outside chance that others would in fact do the same and that, under the circumstances, this act might contribute to a bad state of affairs.

That is to say, the moralizer did not argue that the over-all effects of the one act would be undesirable (or worse than those of some alternative) and that this was the reason against taking some apples. He might have argued in this way while still appealing to utility. But such an argument rests upon applying the test of utility in a radically different way—in what I shall call a *simple utilitarian* way.

Simple utilitarian considerations are those that concern all the effects of the particular act in question (or the effects of that act as compared with those of the alternative acts). If the moralizer had appealed to such considerations he would have asked, 'What will happen if *this* act is performed?' and not 'What would happen if *everyone* did the *same*?'

In contrast, *general* utilitarian considerations concern the total effects that *could* be produced if all acts similar to the one in question, which could be performed, actually were performed. That is, in applying a form of utilitarian generalization, we describe the particular act in some way, thus marking off a class of acts, which could be performed, that are similar in the respects specified. We do not assume that others will do the same. We are only to suppose that the kind of act specified is generally practised and to evaluate the effects of this hypothesized practice.

These two kinds of utilitarianism—simple and general—are distinguishable in two respects: (1) the manner in which value-criteria, the tests for utility, are applied to acts, and (2) the generality of the judgements derivable. In the case of simple utilitarianism, (1) value-criteria are applied to the effects of particular acts taken separately, and (2) judgements concern only particular acts. The rightness or wrongness of a particular act depends upon the value of its effects, i.e. upon its *simple utility*; or alternatively upon the value of its effects as compared with the values of the effects of the alternative acts, i.e. upon its *relative simple utility*. In the case of utilitarian generalization, on the other hand, (1) value-criteria are applied only to what I shall call the *tendency* of an act, i.e. to the effects of everyone's doing the same sort of thing; and (2) the judgements directly derivable concern a class of acts that are similar in the specified way, each one determined as right or wrong or obligatory, or prima facie so, as

the case may be. The rightness or wrongness of a particular act here depends upon the value of its tendency, i.e. upon its *generalized utility*; or alternatively upon the value of its tendency as compared with the values of the tendencies of the alternatives, i.e. upon its *relative generalized utility*.

The generalization test occurs in various familiar linguistic shapes and often incorporates the substance of the matter at hand. Thus we may have:

What if everyone dodged the draft?
Suppose everyone lied just to suit his own convenience?
But suppose everyone failed to pay his taxes!
What would happen if *no one* bothered to vote?

When such objections are made, many kinds of disagreement can arise. For example:

(1) 'Well, what *would* happen? The question you pose is more complex than it may appear. The total effects of everyone's doing what I propose to do may be quite different, qualitatively different, from the effects of this one act. If people haven't generally acted in this way before, we may not know, we may have no reliable idea of, what consequences would result.'

The particular details of the factual problems suggested here will not be considered in this study. We are concerned with the general nature of these principles, and this will lead us to examine some empirical (causal) phenomena. But we shall not be concerned with the specific applications of the principles, and thus not generally with the practical problems of getting the required information and correctly inferring judgements from the principles on the basis of that information.

The practical problems here are akin to, though more complex than, a set of difficulties faced in applying simple utilitarianism. In that case the implications of a given principle depend upon *all* the effects (all the utilities and disutilities) of individual acts, no matter how remote or indirect they may be. Such practical obstacles to success in discovering what a given principle actually implies are compounded in the case of utilitarian generalization, for there one is concerned, not with all the effects of one act, but with all the effects of every one of a class of similar acts, supposing that all are performed.

(2) 'But would the results be as bad as you suggest? Would they be bad at all? How do you judge so? Why in that way?'

This value-theoretic set of problems, in practice linked with (1), will not concern us either. I am distinguishing two features of a teleological or utilitarian theory and dealing with one only. We are leaving value-theory aside and shall concentrate upon the structure of utilitarianism—how the value-criteria are to be applied. Thus we shall not ask 'What are the criteria of intrinsic goodness?' or 'What things are desirable (undesirable)?' or 'How can we decide what is a desirable goal?' We shall consider only questions related to differences in utilitarian theories such as the differences between simple and general utilitarian considerations. We are doing so because some have thought that a mere difference in structure along these lines results in a substantive difference in the implications of utilitarian principles.

But if we do not concern ourselves with value-criteria, and therefore set no restrictions upon them at all, this will allow us to call certain theories 'utilitarian' even though they might not ordinarily be so called. For example, 'self-realizationist' teleological theories might be counted as utilitarian; and is this not a confusion to be avoided? The answer is, that we need not be concerned with such distinctions. It is merely a terminological—and partly historical—point, which principles we choose to call 'utilitarian'. The forms of utilitarian generalization and also the species of rule-utilitarianism that we shall examine are, in fact, usually supposed to be applied in conjunction with universalistic value-criteria (where the interests of each person count equally), and these theories may therefore be counted as 'utilitarian' in one restricted sense. But we are not assuming that a utilitarian theory is necessarily hedonistic, for example (i.e. based upon a pleasure principle), and we need impose no other evaluative restrictions.

The reason some have been concerned to restrict value-criteria used in conjunction with 'utilitarian' principles is that by adopting certain *ad hoc* valuations the utilitarian seems to escape at least some of the traditional criticisms of his theory. Thus, as we shall see, the 'ideal' utilitarian can claim that just distributions are intrinsically good (and unjust distributions intrinsically evil) and thereby attempt to assimilate justice to utility and in that way

accommodate utilitarianism to a class of criticisms based on appeals to justice. But, as I shall argue in the last chapter, even this move will take the utilitarian only so far and not far enough. For what the utilitarian cannot allow is that some value related to the rightness or wrongness of acts is characteristic of acts of certain kinds, e.g. unfair acts, independently of their effects.

The only condition we must impose is that, when principles are compared, the value-criteria employed in conjunction with them must of course be (whatever else they are) identical. This will tacitly be assumed—for our arguments, as opposed to illustrative examples, will be strictly schematic, requiring no specification of value-criteria.

The following issues will receive attention in the immediately succeeding chapters:

(3) 'What is the force of "everyone" in your objection? Who is to count? Surely not everyone, for not everyone will have occasion to do this kind of act. Shall we consider merely those who will pass by this orchard, or all those who will pass by all similar orchards? Or shall we consider only those who will notice the apples? Or perhaps only those who will be strongly tempted to take some? How do we decide which class to consider? How does one show that a particular method of selection is not arbitrary?'

Similarly:

(4) 'What are we to count as the same sort of action? And how do we decide? Shall we consider "picking apples" or "stealing apples"? Shall we mention that no one is looking or that there will be many left when we have taken some?'

The latter two sets of problems are more fundamental than (1) and (2), for any defensible application of utilitarian generalization presupposes answers to the questions raised.

(5) 'But of course not everyone will do the same. To suppose that they will is to suppose falsely. And to act upon such a false supposition intentionally is to mislead oneself regarding the circumstances—and therefore the effects—of one's act—a very *un*utilitarian thing to do.'

(6) 'But of course few others will do the same. Therefore the

evil will not be produced anyway, regardless of what I do, so my act cannot be wrong.'

(7) 'But of course most others will do the same. Therefore the evil will be produced anyway, regardless of what I do, so my act cannot be wrong.'

(8) 'My act itself will not have bad effects. And I am responsible for my acts alone, not for what others will or might do. Thus there is no utilitarian ground against my acting this way—regardless of what others do.'

These objections involve a set of related misunderstandings regarding utilitarian generalization which none the less suggest real problems as to the relevance of the behaviour of others. We shall deal with the relevance of others' behaviour in some detail.

(9) 'Granted that this is an act of the kind you specify; but there are also important differences. This is a special case which deserves special consideration (or indulgence).'

And finally:

(10) 'What does it matter? Why should I consider such an objection at all?'

B. *Development of the New Utilitarianism*

Until recently, the notion of *generalization* in ethics was not normally associated with *utility*. Generalization has had two primary associations: the principle of generality and Kant's ethical doctrines.

The principle of generality—otherwise called, e.g. the principle of impartiality or equity—merely asserts that moral considerations have a universal character or 'bindingness'. A common formula for this notion is 'Treat like cases alike'—and, as we understand, 'Treat relevantly different cases differently.' More particularly, we may say: if it is right (or wrong) for someone to do a certain kind of thing, then it is likewise right (or wrong) for anyone to do a similar thing. Sometimes, the principle is understood as requiring that moral criticism and justification turn upon rules and principles—or at least turn upon general reasons.

The principle of generality thus has a minimal content. It says nothing about *which* acts are right or wrong, nor *why* some are

right and others wrong; nothing about which are to be regarded as similar and which as different, nor why they may be so regarded. In this sense, it is a formal principle (thus, sometimes called the formal principle of justice): it tells us about morality, about the generality of moral considerations, but not about their content, nothing about the rightness or wrongness of acts as such. One might say that it concerns the correctness or soundness of moral reasoning as distinct from the direct assessment of acts (or of other moral subjects).

The Kantian notion of generalization (or of universalizability) is not, on the other hand, strictly formal. Kant's theory directly concerns the assessment of moral subjects (in this case, the maxims of actions, e.g. 'Do such and such' or 'In such and such circumstances, do so and so'); it concerns the application of first-order moral terms such as 'good', and it presumably provides a ground or criterion for their ascription. The test is for the universalizability of a maxim, i.e. whether the maxim of one's act could possibly become a 'universal law' and whether an ideal rational agent could consistently will that it become a universal law.

Clearly, neither notion of generalization is at all related to utility. It would therefore be misleading to speak of utilitarian generalization as 'Kantian' simply because it involves a notion of generalization.

The earliest, pioneering study of a form of utilitarian generalization was made by C. D. Broad two generations ago. ('On the Function of False Hypotheses in Ethics', *International Journal of Ethics*, xxvi (April 1916), 377–97.) Broad considered the more common negative form, that which concerns only the *un*desirable effects of general practices. He pointed out that such a principle is normally applied when it is as certain as possible that not everyone will do the same—that the general practice will not actually occur. The supposition (hypothesis) that everyone will do the same is normally counterfactual. Thus Broad called arguments based on the generalization test 'the method of false hypothesis' or of 'false universalisation' in ethics. He argued that such a principle when viewed as strictly utilitarian was paradoxically most *un*utilitarian, since it required acting upon a 'false account of the circumstances'. This led him towards an 'ideal' utilitarian position. We shall examine these arguments in their turn.

Broad's study neither represented nor occasioned a movement in moral philosophy. It developed as a critique of one common method of moral reasoning which presented obvious difficulties. It is understandable, then, that for two decades after Broad's unfavourable review utilitarian generalization was largely ignored by academic philosophers.

Meanwhile, utilitarianism came under severe attack. Actually, criticism was directed at simple utilitarianism—or at the predominant form of it, *Act-Utilitarianism*. For at that time the distinction between simple and general utilitarianism was unnoticed, and thus Act-Utilitarianism was taken as the paradigm theory.

Roughly speaking (as I shall explain), Act-Utilitarianism is the theory that one should always perform acts the effects of which would be at least as good as those of any alternative. These are right actions; all others are wrong. It is one's duty, or over-all obligation, to perform right acts only; and thus if one act has the best consequences, that act is *the* thing to be done. In our terminology, this grounds the moral assessment of acts upon their relative simple utilities. This theory has otherwise been called 'crude', 'extreme', or 'direct' utilitarianism—although these terms have also been used rather generically with regard to simple utilitarianism.

Admittedly, the classical utilitarian theories might not properly be characterized as purely simple utilitarian. None the less, partly through the influence of G. E. Moore (*Principia Ethica* and *Ethics*), in this century the traditional variety had come to be viewed as simple utilitarian, and Act-Utilitarianism as a coherent formulation of the predominant traditional theory.

Admittedly, there were differences among utilitarians during the first three or four decades of this century—concerning value-theory (e.g. hedonistic versus 'ideal' utilitarians); concerning the scope of moral considerations (positive versus negative utilitarians); concerning responsibility (whether actual or probable consequences should be considered); and so on—but these differences developed within the confines of simple utilitarianism.

Thus, while outside criticisms have mainly been directed against Act-Utilitarianism, their point has been that utility is not the sole (or perhaps not at all a) determinant of right action.

These criticisms have had two related aspects. First, counter-examples were offered, examples of purportedly strong obligations, the existence or strength of which could supposedly not be accounted for by Act-Utilitarians. Criticisms have, for example, turned upon purported 'prima facie obligations' such as those of fidelity, obligations resting more upon past acts or circumstances than upon the effects of present and future acts. It has been claimed that Act-Utilitarianism cannot adequately account for our obligations to keep our promises, to repay our debts, to tell the truth, to punish the guilty and protect the innocent. In particular, it has been held that a really wrong act can appear right on Act-Utilitarian grounds, just because a condition of secrecy shrouds the act. And it is supposed that a condition of secrecy should not weaken our obligations.

In the second place, it has been argued that utilitarians cannot account for certain perfectly good elements of moral reasoning. We often appeal to moral laws or rules or principles—or at the very least to good reasons—in defending particular judgements. We sometimes justify our acts, for example, by saying 'Because I *promised* I would'—or, less typically, by appealing to a principle that promises *ought* to be kept. Such considerations are not obviously utilitarian. How can a utilitarian account for them?

These criticisms may have served to resuscitate philosophic interest in utilitarian generalization. In any event, R. F. Harrod in 1936 offered a 'revised utilitarianism' based upon general acceptance of the main criticisms of Act-Utilitarianism and an attempt to accommodate them. ('Utilitarianism Revised', *Mind*, xlv (April 1936), 137–56.) Instead of scuttling utilitarianism, Harrod sought a new variety. His theory was identical with Act-Utilitarianism in every respect except that the relative generalized rather than the relative simple utility of an act was always to be considered.

Harrod's proposal—including some quite important and original arguments which we shall examine—aroused no immediate interest. This was due, perhaps, to the rise of logical positivism with its associated ethical doctrines: normative ethics appeared to some to be an illegitimate (or at most a psychological) inquiry. Moral philosophers became preoccupied with the nature of ethics and ethical language at the expense of other questions.

Not until after the Second World War did the programme of revitalizing utilitarianism by means of revising it come to be entertained widely and seriously.

The post-war period brought with it the new utilitarianism. This has been called 'modified', 'restricted', and 'indirect' utilitarianism and, increasingly, *rule-utilitarianism*. The principal idea has been to apply the test of utility, not to the effects of an act itself, but rather to its tendency or to a rule under which the act falls.

Sometimes these new terms are applied to particular theories (such as Harrod's); sometimes the terms have, confusingly, been applied generically. It has been confusing because there are different varieties of utilitarian generalization and of rule-utilitarianism, just as there are different varieties of simple utilitarianism. We shall examine some of these intrafamilial differences presently. For the present, let us note that there is some difference and some kinship between utilitarian generalization and rule-utilitarianism. The former makes no direct reference to rules. By rule-utilitarianism I shall mean that kind of theory according to which the rightness or wrongness of particular acts can (or must) be determined by reference to a set of rules having some utilitarian defence, justification, or derivation. Note, however, that particular rules may be assessed by means of a variant generalization test, 'What would happen if everyone observed rule R?', and sets of rules may be evaluated by inquiring, 'What would happen if everyone observed Rules R_1, R_2,..., R_n?'

One source of rule-utilitarianism is the notion of *good reasons* in ethics and the appeal to moral rules or principles. Analyses of moral reasoning have stratified it into several 'levels': first, the justification of particular judgements about the rightness or wrongness of an act by reference to a good reason pro or con kinds of acts, or by reference to a moral rule; secondly, the validation of such reasons or rules by reference to higher-order rules or principles or criteria; and perhaps third (there are variations here), the vindication or ultimate defence of these higher rules, principles, or criteria. (See, e.g., H. Feigl, 'Validation and Vindication', in W. Sellars and J. Hospers, *Readings in Ethical Theory* (New York: Appleton-Century-Crofts, Inc., 1952), pp. 667–80; K. Baier, *The Moral Point of View* (Ithaca, New York:

Cornell University Press, 1958); P. W. Taylor, *Normative Discourse* (Englewood Cliffs, New Jersey: Prentice-Hall, 1961).) Now when the first-order rules (or reasons) are grounded upon a second-order criterion of utility, we have rule-utilitarianism.

To complicate matters, however, the most notable early theories, such as Toulmin's, had impure second-order criteria, not strictly utilitarian. Toulmin placed a special premium upon the social acceptance of rules as opposed to their utilities. (S. E. Toulmin, *An Examination of the Place of Reason in Ethics* (Cambridge, England: Cambridge University Press, 1950).) This impurity in the rule-utilitarian tradition is one reason for coining the new label, 'utilitarian generalization'. But, as I have said, the latter also involves no direct reference to rules.

Another mode of argument intending to lead towards rule-utilitarianism or towards utilitarian generalization is based upon the time-honoured method of appealing to example. By showing that the new type of principle does not fall prey to the traditional criticism, the counter-examples originally offered against utilitarianism in general are vitiated. There is a dialectic involved here which I shall comment on presently. The following examples illustrate one frequent aspect of the argument, that of the revisionistic against the traditional utilitarian:

(1) Should one bother to vote when it is inconvenient to do so? One knows, generally, that his single ballot will not be especially significant; therefore, the direct effects of voting and of not voting will hardly be different, if at all. And, regarding indirect effects, while one's absence from the polls will (let us assume) not be noticed by others, and therefore will not influence their behaviour, it would on the other hand be more convenient not to vote.

If we tally up the score, it appears that an Act-Utilitarian must hold that it would be wrong to vote under such circumstances, since the over-all effects of voting are worse than those of abstaining. Thus it seems that mere inconvenience provides a reason overriding whatever good reason we ordinarily have to vote. And this, many would hold, is simply not so.

An Act-Utilitarian can, of course, object that our facts are mistaken, that we have weighed the utilities incorrectly, that we have overlooked certain pernicious indirect effects. Here

innumerable argumentative complications can arise which, for our present purposes, we need not consider. Let us assume, for the sake of the immediate argument, that there can be such a case of voting that, because of inconvenience, would be wrong on the Act-Utilitarian account.

Now some would hold this as a case against Act-Utilitarianism. It is supposed that one has a good reason for voting that is stronger than Act-Utilitarianism suggests: mere inconvenience (as opposed to serious suffering or hardship) does not provide a sufficient countervailing reason. And, considering the generality of the Act-Utilitarian theory, if the theory is wrong in one case it cannot be accepted; it is not an adequate account of the rightness or wrongness of actions.

The rule-utilitarian—or better, the proponent of utilitarian generalization—might accept these criticisms and yet not reject utilitarianism. He would argue that most voters face a similar predicament, each finding it inconvenient to cast his singly indecisive ballot. If each reasoned in the Act-Utilitarian way he would decide against voting. But what would happen if everyone who found it inconvenient to vote failed to vote? Since, as we are supposing, most will find it inconvenient to vote, very few would then vote; consequently, the wrong man could be elected; or worse still, the mass abstentions might seriously harm the electoral system which is, in the long run, of great importance to all. And this evil would far outweigh the total inconvenience which could be avoided by abstaining. Thus, if *one* abstained, the balance of utility would be positive; no harm would be done, and inconvenience would be avoided. But if *everyone* who found it inconvenient to vote abstained, the over-all effects would be bad, much worse than if all those had voted and suffered the inconveniences. Only when the general practice is considered—and not always when the individual act is considered separately—can we take into account certain undesirable consequences. Only by appealing to utilitarian generalization can certain undesirable consequences be avoided.

But it might be objected that this is illusory. For if others do not act in the condemned way, e.g. if others do vote even when it is inconvenient, the undesirable consequences in question will not materialize whatever one does. And conversely, no

matter what one does, if others do not vote the evil will be produced. The proponent of utilitarian generalization is then pressed to different reasons for urging that utilitarian generalization is none the less a tenable moral principle while Act-Utilitarianism is not. He then argues that a valid moral principle is one that everyone can hold and act upon. If everyone followed Act-Utilitarianism, certain bad consequences would result that could be avoided if everyone in fact acted according to utilitarian generalization. We shall put a fuller perspective upon arguments like these later on, in the third and fourth chapters.

(2) Suppose now that one lives in a town in which racial segregation is the brutally enforced rule, and in which the prospects of changing the oppressive system are quite slim. How should one take these facts into account? One who contravenes the segregation rules endangers his family and friends, jeopardizes his home and livelihood, perhaps removes himself from further activity in the community. For example, fairly certain dangers face a racially mixed group even if they gather at one's home, and face someone who vigorously agitates for reform. The good results of such unorthodox behaviour will (let us suppose) be far outweighed by the vengeful harm done those who refuse to acquiesce in the rules.

It may therefore be argued that Act-Utilitarianism counsels inactivity in such a case—or at least that one refrain from activities which expose one or others to danger.

A proponent of utilitarian generalization might argue, however, that if everyone were to continue to acquiesce, if everyone failed to take the risks entailed by defying and seeking to change the rules, then the suffering imposed by the system would continue unabated. But if, on the other hand, everyone were to run the risks, very desirable consequences would result, far outweighing individual sacrifices which might be made, since the joint effort would be sufficient to change the system.

Again, we have an example suggesting a peculiar divergence between the two kinds of utilitarianism. But the main interest in this case is that personal sacrifice is involved, and that the proponent of utilitarian generalization would hold that, if such sacrifice were called for by his principle, then it must be suffered. Notice, however, that if one or only a few take the risks and

suffer the burdens, that will not be sufficient to produce the good that could be produced by (or to eliminate the evil that could be eliminated by) everyone's doing the same. None the less, it appears that no part of the generalization test concerns what others actually are doing or will do. Thus, these applications of utilitarian generalization make that sort of principle appear quite *un*utilitarian. For the generalizer appears to surrender that hard-headed practicality which has been the hall-mark of utilitarianism. He is not supposed to be concerned with the actual effects of his acts, but only with the conjectured effects of an hypothesized general practice. He may have to accept the sacrifice, in accordance with his principle, on the false supposition that others will do the same.

How should one act? Should one refuse to doff one's hat before the tyrant's statue, knowing that if one acts alone one will suffer as a result and that no appreciable good will come of it, and knowing also that others will not refuse to doff their hats? Are there social situations, political régimes, in which one simply ought not to acquiesce? The relevance of the present cases to these questions is that utilitarian generalization appears to provide a utilitarian ground for disobedience, a utilitarian ground for seemingly *un*-utilitarian but—shall we say?—morally imperative acts. We shall have to examine whether our examples mislead us, whether this is really a utilitarian argument after all. (These issues will arise in the last chapters.)

Let us now consider such examples in the light of the relevant developments in moral philosophy. A three-sided conflict has accompanied the rise of the new utilitarianism. In the first place, critics of utilitarianism have claimed that it cannot account for certain reliable moral beliefs or data. (These arguments, as we have noted, have mainly been directed against simple utilitarianism, although they have also been extended to apply against general utilitarianism. And there have been attacks on the latter by simple utilitarians. However, criticisms of general utilitarianism have either been based on peculiarities of particular theories or presentations of them, or rest upon an inadequate grasp of the nature of utilitarian generalization. We can, therefore, ignore this aspect of the debate.)

Secondly, simple utilitarians have attempted to reject or

accommodate the criticisms. They may claim that the charges are based upon factual error or evaluative oversight; they may go so far as to claim that the charges are based upon moral error; or they may qualify simple utilitarianism, patching it up to meet the objections. Finally, the proponents of utilitarian generalization (or of rule-utilitarianism) generally accept the traditional criticisms of simple utilitarianism while claiming that these are ineffective against their new theories.

Thus, examples such as those I have outlined are supposed to establish the superiority of the new to the old utilitarianism. But what is our method and what are our criteria for criticizing and comparing alternative moral theories? I shall not go into this general question extensively, but I shall deal (in Chapter IV) with a characteristic argument purporting to show the superiority of the new utilitarianism.

It should be obvious, however, that any attempt to displace simple by general (or rule-) utilitarianism presupposes a positive answer to the following questions: Are there in fact any substantive differences between these theories? If so, between which ones? And are these differences in the requisite directions?

Utilitarians have generally failed to examine these questions. Instead their arguments develop along the lines already suggested. A proponent of the new utilitarianism accepts, as (most probably) correct, judgements or generalizations, about particular acts or kinds of acts, that seem inconsistent with simple utilitarianism. Or he accepts, as (most probably) correct, rules or reasons that are used in criticizing or justifying acts but which are prima facie unutilitarian. For example, he may agree with the critics that the strength of one's obligation to keep one's promises (or to tell the truth) is greater than the relative simple utility of promise-keeping (or veracity) would make it appear. In this respect, some hard moral data are more or less assumed—are held less vulnerable to criticism than simple utilitarianism. But it must be observed that the supposed inconsistency of such data with simple utilitarianism is not rigorously substantiated. For any such argument requires that we pin down the facts of the case and specify value-criteria; but some crucial facts are often assumed and the assessment of effects is left at a most intuitive level.

The argument against simple utilitarianism therefore suffers from inconclusiveness. Some such looseness in argument cannot perhaps be avoided, but it is compounded when the revisionistic utilitarian claims—on the basis of sweeping factual assumptions and without specifying value-criteria—that the new theory can indeed account for the data in question.

These comments may appear harsh and unfair, calling for rigour where rigour is impossible. Let us then accept the rough and ready factual and evaluative considerations. One general issue remains: why should we suppose that it makes any difference to assess acts as acts of certain kinds (or as instances of rules) instead of separately? What positive ground do we have—apart from the more or less roughly hewn examples—for making this supposition? This is the issue I shall emphasize: whether in fact the new utilitarianism offers an alternative.

There are certain obstacles to arriving at a firm answer to this question, obstacles arising from the wide variations possible upon the several utilitarian themes. Simple utilitarianism, as we shall see, takes many substantively different forms. Act-Utilitarianism is but one such form—or perhaps it is a genus within the simple utilitarian family, having its own species. As I have mentioned, rule-utilitarianism is often admixed with non-utilitarian elements. Moreover, the reference to rules involves various special conditions depending upon how the rules are characterized. It is impossible, therefore, to make a wholesale comparison between the old and the new utilitarianisms in determining what difference it makes just to structure a utilitarian theory one way rather than the other.

The simplest and most fundamental comparison that can be made is between simple utilitarianism and utilitarian generalization, for these two kinds of principle are defined precisely by reference to such a difference in structure. We shall therefore ask: What difference does it make to apply the test of utility to an act in respect of its generalized instead of its simple utility? What different results are entailed by asking 'What would happen if everyone did the same?' rather than 'What will happen if this act is performed?' And in this way, while directly determining the substantive relations between simple and general utilitarianism, we can begin to sketch the relative position of rule-utilitarianism as well.

c. *The Dimensions of the Principles*

In comparing the two kinds of utilitarianism, we must take care to exclude extraneous factors. Various forms of the two kinds differ substantively, but these differences arise from features with which we are not primarily concerned. For example, Act-Utilitarianism, based unrestrictedly upon the relative simple utilities of acts, differs substantively from those negative versions of utilitarianism which limit our attention to undesirable consequences. On a negative theory, the term 'wrong' can be applied only to those acts that have undesirable consequences on the whole. Thus, according to Act-Utilitarianism, but not according to negative theories, there can be acts with good consequences on the whole which are none the less wrong—so long as their consequences are not as good as those of some alternative.

Here is a difference within the simple utilitarian family itself that could obtrude into our basic comparison. We can eliminate this possible source of confusion by comparing only what I shall call *analogous* principles, i.e. principles that differ only in those respects that distinguish the two kinds of utilitarianism generically.

Consider the following two principles:

(G1) If the consequences of everyone's doing a certain sort of thing would be undesirable, then it would be wrong for anyone to do such a thing.

(S1) If the consequences of a particular act would be undesirable, then it would be wrong for that act to be performed.

These principles are analogous (they form an analogous pair) since they differ only in the relevant respects. They may be viewed in this way:

(G1′) If the generalized utility of an act (viewed as an act of a certain kind) is negative, then every act of that kind is wrong.

(S1′) If the simple utility of an act is negative, then that act is wrong.

It will be helpful now to indicate some respects in which these principles (and any other two analogous principles) are similar. I shall emphasize three: strength, quality, and gradation.

(*a*) *Strength.* A principle is *weak* or *strong* according as it does or does not include the *ceteris paribus* condition, 'other things being equal' or its equivalent. No such qualification is to be found in (G1) or (S1); they are both strong, having this truth-functional structure:

If *p* then *r*.

Here '*r*' is a judgement concerning the particular act or the kind of act in question. For the derivation of such a judgement (i.e. its justification on the basis of such a principle), it is sufficient that the utilitarian condition, '*p*', be satisfied.

These principles differ from otherwise similar but weak forms that yield merely prima facie judgements against acts when the same respective conditions are fulfilled. Thus, the weakened version of (S1) is:

(S2) If the consequences of a particular act would be undesirable, then it would be wrong, other things being equal, for that act to be performed.

In (S2), satisfaction of the initial clause is not a sufficient wrong-making condition, but merely one part of a sufficient condition. In other words, whereas in (S1) the negative simple utility of an act is a sufficient condition for the wrongness of that act, in (S2) it is not sufficient.

This distinction may be viewed in still another way. A principle is weak or strong according as it provides prima facie (i.e. good but not necessarily sufficient) reasons or conclusive (i.e. sufficient) reasons for or against taking or refraining from courses of action. A *good reasons* equivalent of (S2) is:

(S2′) If the consequences of a particular act would be undesirable, then there is a *good* reason against that act being performed.

And (S1) would be rendered, by extending the good reasons idiom:

(S1′) If the consequences of a particular act would be undesirable, then there is a *conclusive* reason against that act being performed.

There are also strong and weak forms of utilitarian generalization (as, indeed, there may be for any kind of principle). One particularly important example is the weakened version of (G1):

(G2) If the consequences of everyone's doing a certain sort of thing would be undesirable, then it would be wrong for anyone to do such a thing without a reason or justification.

This is essentially Marcus Singer's 'generalization argument', which is, of course, merely one of a number of forms of utilitarian generalization. (M. G. Singer, *Generalization in Ethics* (New York: Alfred A. Knopf, 1961).) The qualification 'without a reason or justification' has been preserved from Singer's formulation, although some of the rest is slightly revised into our terminology. It should be observed that this qualification is the equivalent of 'other things being equal' in weakening the principle and thereby limiting the judgements derivable.

While the formulation of a *ceteris paribus* condition is immaterial, of rather more interest is the effect of such a clause upon the truth-functional structure of principles. Weak principles may be analysed in the following manner:

If p and q, then r.

That is, the *ceteris paribus* condition may be viewed as part of the antecedent, as a second condition which must be satisfied if the unqualified judgement 'r' is to be derivable or justifiable. For if only the main utilitarian condition 'p' is satisfied, then the sort of judgement derivable is a weakened or conditional one. Thus, in applying the strong (G1), we may find that the generalized utility of lying just to suit one's convenience is negative. This satisfies the sole condition, and is thereby sufficient for deriving the judgement, 'It would be wrong for anyone to do such a thing (i.e. lie just to suit one's convenience).' But in applying the weak (G2) in the same case, if we merely satisfy the same condition without certifying that other things *are* equal, then we cannot derive the unqualified judgement; we are entitled to derive merely this conditional one: 'If other things are equal, then it would be wrong for anyone to do such a thing.'

(If we view *ceteris paribus* principles in this way, we find that 'prima facie right' and 'prima facie wrong' are definable in terms of 'right' and 'wrong' respectively. And this suggests that the

notion of a 'good reason' can be defined in terms of 'conclusive reason'. I am not sure that theorists who employ concepts like 'prima facie right' and 'good reason' will find this result attractive. Moreover, inserting the *ceteris paribus* condition in the antecedent is to some extent unidiomatic. Indeed, this is but the beginning of a catalogue of potential problems. In general, an analysis of any of these principles—strong or weak—in truth-functional terms may be overly simple and inadequate. Also, there may be grounds for questioning whether there is a radical difference between the sense of the *ceteris paribus* condition in principles and its sense in derived judgements. But I offer this tentative outline of a sketch of *ceteris paribus* principles and judgements in the hope that difficulties merely touched upon here will attract the attention of others. I do not believe that my presentation affects the main arguments to be advanced below.)

The forms of utilitarian generalization can also be cast in the good reasons mould. For example, (G2) becomes:

(G2′) If the consequences of everyone's doing a certain sort of thing would be undesirable, then everyone has a good reason against doing such a thing.

We may note, finally, that the difference between strong and weak principles is substantive. Two principles identical in every respect but that one has and the other lacks a *ceteris paribus* condition have different particular implications. Let us suppose that the main utilitarian condition '*p*' is satisfied. Then an unconditional judgement is derivable from the strong principle, and this judgement simply cannot be overridden. That is, if we assume the correctness of the strong principle, we cannot admit any conflict between the strong judgements derivable therefrom and other judgements. If the conflicting judgements are both strong, we have moral incompatibility. And unless moral reasoning is essentially incoherent, this is an unacceptable state of affairs; one principle or the other (assuming the two conflicting judgements are derived from different principles) *must* be incorrect. On the other hand, if one of the conflicting judgements is weak (derived from a weak principle), it simply carries no weight against a strong, conclusive judgement. Accordingly, if our original principle was weak, the judgements derivable are

weak, and therefore they are simply compatible with any other judgements. They are overridden by conflicting strong judgements, and they must somehow be weighed against conflicting weak judgements. This is a substantive difference. And this is another topic to which we shall return.

(*b*) *Quality*. A principle is *positive* or *negative* according as its application does or does not admit our taking into account positive good that could be produced as well as evil that could be avoided. Negative principles are restrictive: they limit our attention to overall consequences that are undesirable—pain, hardship, suffering, frustration, inconvenience—and require us to don blinders against the utility of acts above some assumed norm or level. This level may sometimes be obscure, but the general distinction seems sound. It is a distinction strongly urged by the proponents of negative utilitarianism.

We may be tempted to render positive and negative principles symmetrically. It may seem, for example, that the positive counterpart of (G1) is:

> If the consequences of everyone's doing a certain sort of thing would be *desirable*, then it would be wrong for anyone to *fail to do* such a thing.

But this would inadequately capture the import of a positive principle. For just as negative principles are intended to be restrictive, so positive principles must be inclusive, comprehending losses as well as gains in utility by treating one as the opposite of the other. (Utilitarians agree that evil should be avoided; but negative utilitarians would restrict the scope of relevant consequential considerations. Thus, we should formulate the positive counterpart of (G1) as:

(G3) If the consequences of everyone's doing a certain sort of thing would be desirable, then it would be wrong for anyone to fail to do such a thing; and if the consequences of everyone's doing a certain sort of thing would be undesirable, then it would be wrong for anyone to do such a thing.

As I have already argued, the distinction between positive and negative principles has substantive ramifications. Thus, the first half of (G3) is not redundant—it has implications apart from those of (G1).

One possible misapprehension about the quality of a principle should be mentioned. It should not be supposed that positive principles yield only 'positive' judgements, i.e. judgements for acts, and that negative principles yield only 'negative' judgements, i.e. judgements against acts. Since negative principles are the more restrictive of the two kinds, the point can most sharply be made with respect to them. Negative principles can yield judgements for or against acts, depending upon such factors as how the act is described. For example, if the results of everyone's refraining from voting would be *un*desirable, then on the basis of (G1) we can infer that it would be wrong for anyone to refrain from voting; and this means that everyone should vote, that everyone has a conclusive reason *for* voting. The nature of the verdict depends as much upon the formulation of the issue (the description of the act) as upon the facts of the case and which principle is applied. This example suggests how we shall interpret the variant generalization test, 'What would happen if *no one* did that?'—employing the 'right'-'wrong' and act-alternative dichotomies. But more on that later.

(*c*) *Gradation.* A principle is *comparative* or *non-comparative* according as it does or does not incorporate some requirement of comparing the utility of an act with the utilities of its alternatives. That is, a comparative principle concerns relative utilities, whereas a non-comparative principles does not. The former involves taking into account differences in degree as well as differences in the quality of utilities.

In the case of strong utilitarian principles, the lack of a comparative feature entails significant defects; in applying such a principle under certain circumstances, anomalous results can be gotten. For example, the strong, negative, non-comparative (S1) provides a conclusive reason against performing any act that has undesirable effects on the whole—it would classify any such act as (simply and unequivocally) wrong. It sometimes does happen, however, that one's alternatives are limited to acts each of which has undesirable effects on the whole. Ordinarily, we would say that one ought to choose the least evil. But (S1) implies that any such act would be wrong. Not in this or that respect wrong, or tending to be wrong, or prima facie wrong—but *wrong*. But not every alternative can be wrong.

Such a case can be set right *ad hoc* if we regard the undesirable effects below a certain degree as unavoidable and therefore not to be counted—thus making at least one act have an indifferent utility. This would avoid the immediate difficulty. The degree would change from case to case. But then we would require another *ad hoc* ruling to cover the application of strong, positive, non-comparative principles in cases where more than one act has desirable effects on the whole—the other edge of the sword. And of course such problems would arise in connexion with utilitarian generalization, concerning the tendencies of acts.

We should therefore want to append to various principles a condition to the following effect: 'unless there is no alternative the consequences of which (or: the consequences of the general performance of which) would be less undesirable (or: more desirable) than those of the act (or: of the kind of act) in question'. Such a condition would transform a non-comparative into a comparative principle.

For our purposes, however, we need not carry along this cumbersome clause. I shall illustratively refer mainly to the two particular forms of utilitarian generalization that have acquired most importance. These two principles are the weak, negative, non-comparative (G2)—Singer's 'generalization argument'—and a strong, positive, comparative form—roughly, Jonathan Harrison's 'modified utilitarianism' and R. F. Harrod's 'revised utilitarianism'. (See J. Harrison, 'Utilitarianism, Universalisation, and Our Duty to Be Just', *Proceedings of the Aristotelian Society*, liii (1952–3), 105–34.)

The latter principle is roughly (G3) plus a comparative qualification, which might be formulated in this way:

(G4) If the consequences of everyone's doing a certain sort of thing would be better (i.e. more desirable) than the consequences of everyone's doing each of the alternatives, then it would be wrong for anyone to fail to do such a thing; and if the consequences of everyone's doing a certain sort of thing would be worse (i.e. less desirable) than the consequences of everyone's doing some alternative, then it would be wrong for anyone to do such a thing.

This is excessively complex. Neither clause in fact stresses the

quality of consequences; what is at stake is the relative generalized utilities of acts. Given the comparative feature and the exhaustiveness and mutual exclusiveness of 'right' and 'wrong', the first clause becomes otiose.

According to this principle, an act is wrong if its generalized utility is less than that of some alternative. This serves to pick out, from every set of alternatives, a class of right acts. There may be several such acts, with equally desirable tendencies, no alternative of which has a better tendency; or there will be one act the tendency of which is better than any other of the original set. The principle can therefore be compressed into a simpler but full formulation as follows:

(G4) If the consequences of everyone's doing a certain sort of thing would be worse than those of some alternative, then it would be wrong for anyone to do such a thing.

The analogous form of simple utilitarianism is the reason for the importance of (G4); it may accordingly be formulated:

(S4) If the consequences of a particular act would be worse than those of some alternative, then it would be wrong for that act to be performed.

According to this principle, an act is wrong if its simple utility is less than that of some alternative. Whatever else holds for (G4) holds, *mutatis mutandis*, for (S4).

This simple utilitarian principle will be recognized as roughly that of Act-Utilitarianism. The first difference between this principle and that of Act-Utilitarianism is the former's compatibility with certain non-utilitarian considerations that are not compatible with Act-Utilitarianism. For the latter is a complete and homogeneous theory in the sense that it admits the relative simple utilitarian considerations of (S4) but *only* those considerations. (S4) on the other hand could be part of a heterogeneous theory which, for example, provided moral grounds for choosing among right acts, i.e. among alternative acts with the highest equally valuable effects. Therefore, to achieve our first approximation of Act-Utilitarianism, we should replace the initial 'If' in (S4) with 'If and only if'. I shall refer to the resulting principle as (AU).

This sort of modification can be imposed upon any other simple or general utilitarian form. If (G_4) were so modified, we would have a homogeneous relative general utilitarian theory of obligation strictly analogous to Act-Utilitarianism. This will be referred to as (GU).

Other modifications might be made in this resultant Act-Utilitarian principle. The most important of these are the probability qualifications, e.g. where an act is assessed according to its probable instead of its actual effects. I shall comment on these briefly.

I have noted three dimensions of utilitarian principles: strength, quality, and gradation. Within these dimensions as I have sketched them, we can construct eight analogous pairs of principles—or sixteen, if we allow for the strengthening transformation based upon changing 'If' to 'If and only if'. It is possible, however, to make finer distinctions. One type is within the probability dimension. All the principles we have considered so far take into account the actual effects of acts, as opposed to their probable or expectable or foreseeable or even intended consequences. The probabilistic variants will not be considered here, for several reasons.

In the first place, it is difficult to formulate a probability qualification, partly because there are alternative approaches to its conception (especially when intentions come in), partly because there are various ways of calculating probabilities. Secondly, it would be extremely difficult to develop some of the arguments I shall hazard below if probabilities were generally considered, if only because the arguments would be that much more complex.

Moreover, one may be unconvinced that probability qualifications are really desirable. What is the point of including such conditions? Often, the avoidance of a morally untenable position. Since the effects of a particular act are so complex and far-reaching, with perhaps completely unexpected, unexpectable, not to say unintended ramifications, one cannot generally be expected to know the actual effects of acts. A man must therefore generally govern his conduct on the basis of what is most likely to happen (or what can reasonably be expected to happen, or what he intends to accomplish—or what have you). If we tie the

rightness of acts to their actual effects, we aim too high, requiring of a man more than he can reasonably be expected to deliver. It would be unreasonable to claim that Jones ought to have done *x* because *x* had the best effects, when we are aware that Jones could not have known this and that he acted according to his best lights. If we tell him that he should have done *x*, i.e. that it was wrong not to have done *x*, we are telling him that he ought to have done what he could not necessarily do. But we cannot demand that he act upon the basis of the actual effects of acts, for we cannot require that he know (or even feel certain of) their actual effects.

This approach suggests, however, a conflation of two distinguishable kinds of moral considerations. Our subject, on the one hand, is the rightness or wrongness of acts. Another kind of moral consideration—outside the scope of this inquiry—relates to the moral worth of persons: questions of responsibility, of praise and blame, of morally justified expectations regarding others' behaviour, of morally defensible deliberation. The criteria involved in these two kinds of considerations may be very different. While a theory of obligation—concerning the rightness or wrongness of acts—might hold that Jones ought not to have done *x*, i.e. that *x* is a wrong thing to have done, this theory is compatible with a theory about the moral worth of persons which holds that Jones is not blameworthy for having done *x*. The question whether Jones could have known that *x* was—'objectively', or all things considered—a wrong thing to do may have no bearing upon Jones's error in doing *x*.

In any event, our simple, non-probabilistic principles may be qualified, if necessary, with some probability condition. We may view the principles that we shall consider below as raw material admitting further refinement. Accordingly, it may be helpful to regard the succeeding chapters as to this extent not a full study of utilitarian generalization, but a basic sketch.

The second point of divergence between (S_4)—or (AU)—and some approaches to Act-Utilitarianism, concerns the value-criteria which may be employed in conjunction with this kind of principle. We are, as I have said, simply disregarding questions in value-theory proper. This renders our conclusions the more general.

Finally, a note on terminology. I am using a simplified first-order normative vocabulary based upon the terms 'right' and 'wrong'. It is irrelevant here whether these are the central notions in ordinary moral reasoning; they serve quite well, I believe, in conveying the substance of utilitarianism. In this context, I take it that 'right' (in the sense of conclusively right, and not merely right other things equal) is mutually exclusive with 'wrong' (in the sense of conclusively wrong). And these terms are exhaustive of the possibilities—although finer distinctions can be made. That is to say, we are supposing that while acts may be in this respect right and in that respect wrong, any particular act (or any act which falls under the purview of moral considerations of this sort) is *either* right *or* wrong. In other words, two judgements concerning a particular act x, one of which asserts that x is right and the other of which asserts that x is wrong, are morally incompatible. If we are not unwilling to use such terms in ethics, we can say that a theory which implies morally incompatible judgements is inconsistent. Thus, in accordance with modern utilitarian usage, I shall say that 'right' is equivalent to 'not wrong' and 'wrong' to 'not right'. I would not myself argue that every act is in fact either right or wrong, as these terms are normally used. But the usage we shall make of these terms is to some extent technical, in the context of a very general type of normative theory.

Secondly, I take it that judgements are substantively different or non-equivalent only when they cannot be transformed into identical judgements by means of the linguistic moves suggested in the foregoing; otherwise they are substantively identical or equivalent.

Thirdly, at the risk of excessive caution, I shall note that in speaking of *the effects of* an act I intend to include *all* the effects; in speaking of *the tendency of* an act I intend to include *all* the effects resulting from *everyone's* doing acts of the sort specified. In speaking of the *utilities* of an act (simple or generalized), I intend to include *all* value which is teleologically attributable to that act (or to the acts of the kind in question), i.e. the net-balance of desirable or undesirable consequences. This manner of speaking presupposes, of course, that particular acts are discrete and identifiable, and that they can in some way be distinguished from their consequences. This presupposition is problematic, but I do

not see how we can avoid it in working within the framework of utilitarianism.

I shall use the terms 'act' and 'action' indifferently. In general, I shall use lower-case letters, x, y,..., to stand for particular acts, and upper-case letters, A, B, C,..., as place-holders for descriptions of acts (kinds of acts). One or two other special terms will be introduced in their appropriate settings.

And now we can properly formulate the issue involved in our comparison of simple utilitarianism and utilitarian generalization. We shall be concerned to determine whether there can be a condition of extensional non-equivalence (non-equivalence, for short) between analogous forms of the two kinds of principle. Non-equivalence obtains if, and only if, analogous principles do not always yield substantively identical (equivalent) judgements with respect to particular acts.

On one view, the judgements which are directly derivable from the forms of utilitarian generalization are themselves general, relevant to acts of certain kinds. But particular judgements are in turn derivable from such general ones by means of a mediating premiss that a particular act is of the kind specified in the general judgement. In principle, it is these particular judgements ultimately derivable from the forms of utilitarian generalization which are to be substantively compared with the particular judgements directly derivable from the analogous forms of simple utilitarianism.

For example, we can perhaps derive from (G1) this general judgement: 'It would be wrong for anyone to lie just to suit his convenience' (on the supposition that the tendency of lying just to suit one's convenience is undesirable). Noting that the particular act in question is such an act—can properly be so described, we derive a conclusive judgement against the act. And if the analogous simple utilitarian principle, (S1), does not yield an equivalent, a substantively identical, conclusive judgement against the same act, a condition of non-equivalence obtains.

II

DESCRIBING AN ACTION

IN applying utilitarian generalization we view an act *as* an act of a certain kind. But of *which* kind?—for any act admits of innumerable characterizations. What features of acts should we consider? This is the problem of relevance, a solution of which is required before we can hope to apply such a principle.

There is a related but as we shall see more restricted problem, that of inconsistency. It seems that we can generate incompatible judgements from the forms of utilitarian generalization. If so, they cannot reasonably be regarded as sound moral principles.

We shall find that a solution is possible to each of these problems which requires that we take note of the nature or content of the principles—first, as utilitarian, and second, as general utilitarian principles.

A. *Relevance and Consistency*

When a form of utilitarian generalization is applied, our attention is directed to the supposed performance of a class of similar acts. This class is marked off by certain characteristics which each member possesses; every act in the class is similar in a determinate way—no matter how different in other respects—and the class includes every such act.

Consider, first, the role of the term 'everyone' in the forms of utilitarian generalization and in the generalization test itself. Since our general subject is the rightness and wrongness of actions, our concern is presumably limited to the class of moral agents. But when we ask, for example, 'What would happen if everyone failed to vote in the forthcoming elections?' we cannot be concerned with every such person, for not every moral agent will have occasion to do such a thing. We are concerned only with those who might or might not fail to vote. Only one who is a franchised voter can fail to vote if he does not vote, for the option is open only to him; no others can be considered. In

general, then, the effective scope of the term 'everyone' depends upon the given description of the act: it is restricted to that class of persons each of whom will have occasion to do the sort of thing specified, to each of whom such a course of action is or will be a practical possibility.

This class of agents varies according to the description of the act and the facts of the case. Some kinds of actions are open to very few, others are very general options; under some descriptions there is only one whereas under other descriptions there are several opportunities for such an act for each person—depending upon the circumstances. Thus, telling lies when it suits one's convenience is a kind of act that is open to a very large number of individuals, but to whom and how often depends upon the actual circumstances various individuals come to face. In contrast, fulfilling one's responsibility as President of the United States is something which only one person at a time can do or fail to do, and it applies to very few over a long period of time.

The description given may be more or less general, and it may include not only a characterization of what we would ordinarily count as an act (or action), but also a specification of the agent and of surrounding circumstances as well. The generalization test may therefore be understood in the following way: 'What would happen if every *similar* person did a *similar* thing under *similar* circumstances?' For simplicity's sake, however, we shall give the notion of an *act* (or of a kind of act) a special sense. When we speak hereafter of the description of an act it will be assumed that this includes everything that is mentioned about the agent and his circumstances as well. This will allow us to speak of acts which are similar in all relevant respects without the fear that a particular class under consideration might be confused with somewhat similar acts which are performed by relevantly different agents or in relevantly different circumstances. And this convention will allow us to continue to regard the generalization test as 'What would happen if everyone did the same [sort of thing]?'

If we are to make any sense of the generalization test, whenever it is employed the class of acts to be considered must be made determinate. Sometimes the schematic form of the test is used, and the intended description is not explicitly given, but must be

inferred from the context of its occurrence. If the description is not evident, of course, one can request it and expect a response. On other occasions, the expression 'the same' is replaced by a description of the particular act in question. For example, instead of saying 'Suppose everyone did that!' our moralizer might have asked—or might have been led to ask—'What would happen if everyone stole apples?'

Thus, a form of utilitarian generalization is applied to a given act in respect of some description, some characterization of that act which marks off a class of similar acts. But any particular act can be described or specified in innumerable ways: an unlimited number of true things can be said about it. Yet it would be impossible to consider—indeed to list—all true or correct descriptions of a given act. Moreover, even if such a listing were possible, it would seem to defeat the purpose of utilitarian generalization to consider the entire list, for such a description would mark off a class composed solely of that act. And the point is to specify certain features of the given act which can be common to a number of acts. Otherwise the reference to 'everyone' would in every case be vacuous. Furthermore, every act is similar to all others in one respect or another; and yet each act is different from every other in some respect. Some selection of descriptions must therefore be made. Our problem is, to begin with, *what* selection is to be made—or, *how* such a selection can reasonably be made.

One might think that the selection of descriptions, the actual specification, is up to him who employs the test or applies such a principle. Indeed, with respect to an initial specification, this must in a sense be so. But any specification is selective, and this gives rise to the question, whether there are grounds for the assessment of descriptions, whether some descriptions are better or more adequate than others.

For example, any given description may omit what appear to be significant features of the act in question. One might say that he is merely picking apples (and what difference if everyone did that?), while another might point out that he is picking them without the owner's permission. In return, it might be indicated that there will be many left when these few are taken; that there are no witnesses; or perhaps that one is starving and that there is no other recourse. All these things may be true of the one

particular act. Moreover, the act may be regarded as a case of stealing and of trespass—both violations of law; as an act performed on Tuesday at three o'clock; and so on. (The various partial descriptions need not be directly related, though they can always be combined more or less artificially as species and genus.) If all these things (and many more) are true of the one act, which of them are to be considered? Are some particularly relevant and others not? Where is the line to be drawn? Is there a criterion of relevance?

With a criterion of relevance we could select some descriptions and ignore or reject others. We could then, if we wished, apply a form of utilitarian generalization to the act in a single determinate respect, taking into account all the relevant features of that act (including the agent and circumstances) in one description. This might, of course, be an ideal which we could only roughly approximate in practice; but we might *try* to apply the principles in this way and then consider any application *defeasible* in this sense: if we found that we had included an irrelevant feature of the act or that we had ignored some relevant feature, we could make the appropriate adjustments, derive a new judgement, and disregard the judgement previously derived on the basis of the inadequate description.

If we do not apply these principles to acts in such a way, with respect to one description only in each case, there will be something odd about our speaking of *the tendency of* an act or of *the* consequences of everyone's doing *the same*. For, speaking in such terms, we seem to imply that one determinate specification is to be considered with which is associated one particular tendency. What counts as everyone's doing *the same* depends on what description we choose; and what consequences are accordingly to be considered depends on what class of acts are supposed performed. This oddity may however be acceptable; in any event it is not a decisive factor in favour of the method of application just suggested.

Alternatively, we might adopt a method of applying the forms of utilitarian generalization that allowed more than one acceptable description and more than one application of a given principle to a given act. A principle would be applied over and over again, each time in respect of a different description of the act. Such a method will be examined presently.

It might also be suggested, since we are dealing with purportedly moral principles, that we are (or should be) concerned only with those properties of acts, with those similarities among them, that are of *moral* relevance. But how are we to determine which properties of actions are, in general, morally relevant and which are not? We cannot simply say, for example, that surely it makes no moral difference whether the act is done at two or at three o'clock; nor that surely we must consider that others' interests are being trampled. But even if we could confidently say these things, what should we say about the relevance of this or that motive, or about the relevance of the fact that no witnesses are present? Clearly, any justifiable ascription of 'relevance' or 'irrelevance' to a particular description or to a part thereof must be based upon a theory of relevance or upon an explicable criterion. This is what we are seeking.

In the second place, it would be erroneous to view our quest as one for a criterion of moral relevance *in general*, or for a criterion of relevance for the application of utilitarian generalization which must be morally acceptable. Suppose we could, somehow, arrive at a general criterion of moral relevance for the description of an action. Suppose also, however, that we can determine a criterion of relevance which is peculiarly suited to the application of a form of utilitarian generalization. These two criteria need not be identical; they may conflict. If we wish to understand utilitarian generalization, to grasp its peculiar import, we should apply such a principle with a criterion of relevance which is implied by the principle itself, whether or not such a criterion agrees with our notions of 'moral relevance'. I shall argue, in the following, that we can discover such a criterion, and I shall provide a sketch of what it involves.

Let me suggest now an approach to the problem of relevance which will also bring out the second general problem, that of inconsistency. Each different description marks off a different—or at least differently determined—class of acts. One act, through its various descriptions, is thus associated with a number of different tendencies. Suppose that act x is A, B, C, and so on. When we generalize, when we consider everyone's doing *the same*, we do not consider everyone's doing x's, but rather everyone's doing A, or B, or AB, or AC, and so on. And the results of

everyone's doing *A* may very well be different from—and different in value from—everyone's doing *B*, or *AB*, or *AC*, and so on. Thus, one act may be said to have, not only a number of tendencies, but tendencies of different value, and therefore different generalized utilities. This sort of complication does not arise in connexion with simple utilitarianism, for the simple utility of an act is not description-relative.

For example, let *x* be an instance of picking apples, among other things. We may be unsure of the generalized utility of picking apples (without qualification, i.e. whenever such an occasion arises), but we can be quite certain that the several species of this genus-specification will differ in value. The tendency of picking someone else's apples without his permission will undoubtedly be less desirable, perhaps undesirable on the whole. But the tendency of picking someone else's apples without his permission, when one is starving and there is no other recourse—which might also be a description of *x*—may very well not be undesirable at all.

Thus, adding or subtracting properties of an act can make a great deal of difference to the results of applying utilitarian generalization. For first, a difference in the specification of an act can make a difference to its generalized utility; and second, when a form of utilitarian generalization is applied, the rightness or wrongness of an act depends solely upon its generalized utility.

These observations suggest a starting-point towards solving the problem of relevance. For a difference in the generalized utility of an act that is associated with a difference in its description seems to signal a relevant difference in description. Since 'belonging to someone else, and without his permission', when added to 'picking apples', modifies the generalized utility of the act, the addition may be presumed to be relevant—at least for the application of *this* kind of principle.

Let us consider a case in which a form of utilitarian generalization is applied more than once to a given act in respect of different descriptions. This method of applying such principles I shall call *the method of rebuttals*, a method which will be discussed more fully in the succeeding section. For the present, it will serve to introduce the second problem in the application of utilitarian generalization.

Consider the classic case of someone lying in order to save another's life. A third party (perhaps the misdirected axe-wielder) might object to the act by demanding, 'What if everyone lied?' and by invoking:

(G1) If the consequences of everyone's doing a certain sort of thing would be undesirable, then it would be wrong for anyone to do such a thing.

Now suppose that the tendency of lying is undesirable, i.e. if everyone were to lie (whenever the occasion arose), the consequences would be undesirable. If we viewed the act as an instance of lying then we could infer from (G1) that this particular act, *y*, is wrong.

But someone might attempt to rebut this. The act is not merely a case of lying, since it is done in order to save another's life. And it might be argued, first, that if everyone lied under these circumstances (i.e. whenever such a special occasion arose), the consequences flowing from these acts would not be undesirable. Hence, *y*, which is such an act, cannot be wrong. Alternatively (and in this case much more soundly, as we shall see), one could argue that it would be wrong not to do *y* because the consequences of no one's lying under these circumstances would be undesirable.

I shall offer a critique of these arguments and counter-arguments below. For now, let us observe the nature of the last rebuttal. This yields, not actually a rebuttal of the original argument, but a judgement which is incompatible with it. Both the argument and counter-argument are developed from the same principle, (G1); both apply to one act (though of course the act is viewed in different ways). And if we assume that an act is wrong if and only if it is not wrong not to do it, we can see that these two judgements are mutually inconsistent.

Thus, in presenting a theory of relevance and a method of applying the forms of utilitarian generalization, we must take cognizance of two facts. First, it can make a great deal of difference how an act is described. Secondly, if (at least some of) these principles are applied in a certain way, they will yield incompatible judgements.

B. *The Method of Rebuttals*

On one view of moral reasoning, principles of conduct are conceived as providing reasons pro and con an act, reasons which must somehow be weighed. This approach gains credence from our common experience of having (or at least thinking we have) good reasons, often conflicting, sometimes extremely difficult to reconcile, which pull us this way and that when we must decide upon a course of action. Such reasons as we actually consider are not, of course, of one kind: we consider our interests and aims, and we often consider those of others; we consider what would be just or fair, what our duties and responsibilities are, the expectations others may have of us, and so on. If we can abstract from these certain peculiarly *moral* considerations, we may view some of the conflicts as between different moral principles. For example, the demands of justice sometimes seem to oppose those of interest—and the two do not always seem reconcilable.

Thus, in constructing a logical (as opposed to a psychological) analysis of moral reasoning, we may view conflicting moral considerations as conflicts between the implications or demands of different moral principles. And these conflicting implications or judgements may be viewed as if presumptions for or against an act based upon one principle can be countered by rebuttals based upon another principle.

The method of rebuttals is analogous to this approach to moral reasoning. It is based upon the fact that sometimes when we apply a particular moral principle in practice we are unable to take all relevant factors into account. And our failure to consider some factors results in our incorrectly inferring a judgement for or against an act from the principle. Thus we want in practice to regard the judgements that we think are implied by the principle as tentative, subject to correction in case something significant proves to have been overlooked. We do not want to say that a simple utilitarian principle actually implies that an act is wrong just because the act has bad consequences of which we are aware, when we are not aware of the overriding good consequences. Similarly, certain relevant features of the description of an act may be overlooked (or irrelevant features mistakenly included) when we come in practice to apply a form of utilitarian generalization. And

we want to regard such inferences as tentative. This approach, extending the dialectical model of presumptions and rebuttals to the application of a *single* principle, is the method of rebuttals.

The following is a general outline of the method as it applies to utilitarian generalization. (1) There is some minimal test or condition of relevance which any description or part thereof must pass or satisfy if it is to be used in a judgement. (2) We assume that there are, or can be, a number of acceptable descriptions of a given act with respect to which a particular principle can properly be applied. Any given description that passes the minimal relevance test might be shown to be inadequate in some respect. (3) The general rule is, that other things equal the more complete description is the more adequate or more acceptable one. Thus, descriptions can be graded as more or less acceptable on the basis of relative completeness, and this is the basis upon which judgements can be weighed. A judgement based upon a more complete description takes precedence over one based on a less complete description. I shall illustrate this for one form of utilitarian generalization.

First let us consider the condition of relevance which is part of the method of rebuttals. The main idea is that acceptable descriptions are ones that can be employed in judgements derived from a given principle. A certain range of generalized utilities can be said to be significant for a given principle: these are generalized utilities that satisfy the main antecedent condition of a form of utilitarian generalization. In this connexion we need be concerned with only two dimensions of a given form: with its quality and gradation. (The strength of the principle affects only the strength of the judgements derivable.) The property mentioned in an acceptable description must be such that its associated generalized utility is significant for a given principle.

Suppose, for example, we are applying a negative non-comparative principle. An acceptable description for the application of such a principle must have an undesirable tendency (a negative generalized utility). It must be the case that, if everyone did *that* sort of thing (acts of the kind mentioned), the consequences would be undesirable on the whole. Thus, if the description 'lying' has a tendency which is undesirable, then such

a description is acceptable on this approach. For a judgement can be derived from a negative non-comparative principle in such a case. Thus, if an act happens to be a lie, a judgement against such an act is forthcoming. But it must also be noted—this will prove important later—that descriptions which do *not* have *un*desirable tendencies cannot be acceptable; for a negative principle cannot yield any judgements at all about acts in respect of tendencies that are not undesirable.

Accordingly, since judgements can be derived from a positive principle in respect of desirable as well as undesirable tendencies, descriptions which are related to positive as well as negative generalized utilities are acceptable in such cases. And so on.

(Readers familiar with *Generalization in Ethics* will recognize that there is much in common between what I am calling the method of rebuttals and Singer's method of applying—and finding acceptable descriptions for the application of—his 'generalization argument', our (G2). My concern here is not to present an analysis of Singer's method, however, but to present a full and coherent account of a method along those lines. Singer's presentation is, it seems, neither full nor coherent. I shall indicate some difficulties below, mainly for the sake of comparison and cross-reference.)

Let us now reconsider, in the light of these general comments, the previous example of someone lying in order to save another's life.

(i) *Direct argument using* (G1). The act in question, y, is A (i.e. y can correctly be described or characterized as A; 'A' is true of y). Now let us suppose that the tendency relative to this description is undesirable. For example, 'A' might be 'lying'. Since (G1) is a negative and non-comparative principle, the undesirable tendency of A satisfies the antecedent condition of (G1). Thus from (G1) we can infer that every A is wrong; and, since y is A, that y is wrong. Moreover, since (G1) is strong, the judgement against y is conclusive.

(ii) *Direct counter-argument using* (G1). The act in question, y, is also B. Here 'B' might be 'performed on a Tuesday' or 'lying in order to save another's life'. Now if the tendency associated with B is *not* undesirable, we may be tempted to say that, according to (G1), it is not the case that every B is wrong; that, since y is B,

y is not wrong. But this counter-argument is fallacious. And it is important to see in it the following defect: (G1) implies nothing about *B* in respect of its having a tendency that is *not* undesirable; and it therefore implies nothing whatsoever about *y* in so far as it is a *B*. So it is simply impossible to rebut the original presumption against *y* in this manner—by reference to *y* being a *B* and *B* not having an undesirable tendency by appeal to (G1).

Of course, it is also true that one cannot condemn *y* on the basis of its being a *B*; no judgements against *y* in respect of its being a *B* are derivable from (G1). But on the other hand, no judgements for *y* in respect of its being a *B* are derivable from (G1). One is simply unable to derive a judgement about *y*. Thus, this counter-argument fails, and the original presumption remains, totally unaffected, since it turns upon *y* as *A*. (It seems that Singer makes the mistake of supposing that a direct counter-argument is possible using (G2); see Singer, pp. 141–5.)

A digression is in order. Note that *B* is therefore not an acceptable description of *y* for the application of (G1). It makes no difference what relation *B* has to *A*. As I mentioned, '*B*' might be 'performed on a Tuesday' or 'lying in order to save another's life'. In the latter case, although the description '*B*' is more complete than '*A*', than 'lying', '*B*' is not acceptable, so it is not more adequate. For '*B*' does not satisfy the condition of relevance: it cannot be employed in the derivation of a judgement from (G1). All this holds as well for (G2), the weak counterpart of (G1).

But compare the principle we would get from (G1) simply by adding 'and only if' after the initial 'If'. This strengthening would enable us to develop a direct counter-argument on the basis of a tendency that is not undesirable. Whereas an undesirable tendency is, in the case of (G1), a sufficient wrong-making condition, it is not a necessary one; the addition of 'and only if' makes an undesirable tendency both a necessary and sufficient wrong-making condition. In the case of the weak (G2), of course, an undesirable tendency is, while not a sufficient wrong-making condition, none the less a sufficient condition for *presuming* an act wrong. But if we add 'and only if' similarly to (G2), then an undesirable tendency is for the resulting principle both a necessary and sufficient wrong-presuming condition.

Again, such a direct counter-argument could be developed on

the basis of (G3), the positive counterpart of (G1), if the tendency of 'lying in order to save another's life' is desirable and not merely not undesirable, i.e. provided it is not indifferent. Thus, for such a principle, this description would be acceptable. One could continue in this manner to explicate the differences between the various forms of utilitarian generalization with respect to these two related factors: their substantive import and the descriptions of particular acts that, under similar conditions, are acceptable for the application of some but not of other principles.

Now to resume.

(iii) *Indirect counter-argument using* (G1). A plausible counter-argument can be developed indirectly. Here we make appeal to the variant form of the generalization test: 'What would happen if *no one* did that?' We cannot work with the test in this form, for our principles tell us nothing directly about the consequences of no one's doing any sort of thing. But we are, none the less, inclined to say that if the undesirable consequences of a universal performance count against every act of a certain kind, then the undesirable consequences of a null performance likewise count for every act of a certain kind.

We can assimilate the variant test to the standard form. The variant test is applied with respect to some description of the act in question. Suppose *y* is *C*. The indirect counter-argument then develops if the results of no one's doing *C* would be undesirable. For when we ask, 'What if *no one* did *C*?' we are in effect asking, 'What if *everyone* did *not*-*C*?' Here it must be understood that 'not-*C*' is the complement of '*C*'; it is not the description of a particular act, nor merely any description other than '*C*'; rather, doing not-*C* is the exhaustive alternative of doing *C* in the given case. Doing not-*C* is doing anything other than *C* when *C* is a kind of act open to the agent. ('Not-*C*' will be abbreviated hereafter '$\bar{C}$'.)

Thus, if the results of no one's doing *C* would be undesirable, then it follows that the tendency of $\bar{C}$ is undesirable. And therefore, every case of $\bar{C}$, i.e. not doing *C*, or failing to do *C*, or refraining from doing *C* (when *C* is an option), is wrong; hence it would be wrong not to do *C*, in every case; hence *y*, as an instance of *C*, is an act it would be wrong not to do.

Let us say, for example, that 'lying in order to save another's

life' is '*C*'. The complementary description is 'not lying-in-order-to-save-another's-life' (with hyphenation added here only to ensure that the whole and not some part of the description is negated). That is, doing anything other than lying-in-order-to-save-another's-life, when the latter is an option, is condemned by this counter-argument.

However, as already indicated, the original argument and this counter-argument are not properly viewed as merely a presumption and its rebuttal. For these judgements are strong—derived from a strong principle—and they are incompatible. It cannot be both wrong to do *y* and wrong not to do *y*.

It might be objected that the foregoing is a parody on the notion of presumptions and rebuttals, on the notion of good reasons in ethics. Indeed, that is the point of this presentation: to show that a method of rebuttals is inappropriate for the application of strong principles (unless it is arbitrarily restricted to disqualify recalcitrant cases). If we want pros and cons which can be weighed, we do not want to use strong principles. But, on the other hand, if we want to deal with strong principles, this is the wrong method of application. Let us therefore turn to weak forms of utilitarian generalization.

In his defence and advocacy of (G2), Marcus Singer seems to address himself to the problem of incompatible judgements developed from a single principle with respect to a given act. (Singer, pp. 71–83.) He applies the weak, negative, non-comparative (G2) roughly in accordance with a method of rebuttals —except that he adds one major and one minor restriction, the latter supposedly reducible to the former. The major restriction, against so-called 'invertible' applications of (G2), is worth examining for the light it sheds upon the logic of weak principles and the need to take extreme care in explicating their import.

Singer argues:

> Since the consequences of everyone's producing food would be undesirable, on the pattern of the generalization argument [i.e. (G2)] it would seem to follow that it is wrong for anyone to do so, and this, of course, is absurd. (p. 72.)

But Singer does not regard this absurdity as a counter-argument

against (G2). Referring to the judgement against producing food, Singer continues:

> But this actually does not follow, and the generalization argument does not at all have this consequence. For consider what would happen if no one produced food. If no one produced food, everyone would starve. Hence on the same line of reasoning it might be argued that everyone ought to produce food. The argument that no one ought to produce food because of what would happen if everyone did can thus be met by the counter-argument that everyone ought to produce food because of what would happen if no one did. A valid application of the generalization argument, however, cannot be met by such a counter-argument. (p. 72.)

That is to say, a 'valid' application of (G2) cannot be 'inverted' in this way. We may say, in our idiom, that a given description is unacceptable, according to this restriction, if it is the case both that (i) its tendency is undesirable, and (ii) the tendency of its complementary description (here, 'not producing food') is also undesirable.

I shall not consider here whether this restriction against 'invertibility' is—as Singer suggests—somehow 'implicit' in (G2), nor, on the other hand, whether it is—as it seems to be—merely *ad hoc*. Of considerably greater interest are the point of and the supposed need for such a restriction. For the passage suggests that Singer seeks to dispel some spectre of inconsistency hovering about his principle. It would seem, from his presentation in this passage and elsewhere, that such 'invalid' applications of (G2) yield two incompatible judgements such as we have seen can be generated from (G1) by means of the method of rebuttals. In this case, the judgements he rejects are, in terms of 'wrong':

Argument: It would be wrong for anyone to produce food.
Counter-argument: It would be wrong for anyone not to produce food.

But we must be clear about the character of (G2):

(G2) If the consequences of everyone's doing a certain sort of thing would be undesirable *and* other things are equal (i.e. no reason or justification *can* be adduced), then it would be wrong for anyone to do such a thing.

(G2) is a weak principle: the *ceteris paribus* clause of such a principle conditions the judgements derivable in such a way that incompatible judgements are precluded. Let us consider first the judgements that actually are derivable in the present case. All we have here is the satisfaction of the main antecedent condition of (G2). The resulting judgements are, therefore:

Argument: If other things are equal, then it would be wrong for anyone to produce food.
Counter-argument: If other things are equal, then it would be wrong for anyone not to produce food.

And these judgements, as I shall show, are not incompatible.

How might one be led to believe that these two judgements, based upon a given specification and its complement, are incompatible? Clearly one way is to ignore, overlook, or miss the point and force of the *ceteris paribus* condition. But let us suppose that we acknowledge that condition: one might still exaggerate the conflict between these two judgements. For it is easy to misinterpret the *ceteris paribus* condition: the clause retains the same form in each occurrence and it may therefore appear to be a constant. One might then view the conflict in this way, taking '*p*' as the *ceteris paribus* condition:

Argument:	If *p* then *q*
Counter-argument:	If *p* then *r*
Lemma:	If *q* then not-*r*
Therefore:	If *p*, then *r* and not-*r*.

The lemma is an instance of: 'If it would be wrong for anyone to do a certain sort of thing, then it is not the case that it would be wrong for anyone not to do such a thing.'

Consider the result:

'If *p*, then *r* and not-*r*.'

Although this is not itself a truth-functional inconsistency, it has a special character. For if *p*, then there is an anomalous situation. That is to say, if other things *are* equal, then it would be both wrong and not wrong to produce food.

Is this not bad enough, then? There could, presumably, be

at least one case in which other things *are* equal; and in that case such an anomaly would result. Perhaps this contingency is what Singer seeks to avoid.

But the foregoing represents a totally unsatisfactory way of rendering the conflict between two weak judgements. For the *ceteris paribus* condition is no more a constant than 'this' or 'in this respect' is. To render the condition as '*p*' in both judgements is to commit the Fallacy of Equivocation. The *ceteris paribus* condition is a systematically varying clause the specific content of which in any case is linked to the description employed in a given judgement. Instances of the *ceteris paribus* condition are identical if, and only if, the very same description is employed. But if the same description is employed in two judgements derived from a single principle, these judgements will be identical.

Actually, the sense of the two weak judgements is:

Argument: If things other than everyone's producing food are equal, then it would be wrong for anyone to produce food.

Counter-argument: If things other than everyone's *not* producing food are equal, then it would be wrong for anyone *not* to produce food.

And these are clearly not incompatible. Clearly the *ceteris paribus* condition has its point in just this sort of conflict. If other things are equal in one case, then they will not be equal in the other. That is to say, to suppose that both conditions are simultaneously satisfied is to suppose that (i) with respect to producing food, there are no other relevant features of the act in question, and (ii) with respect to *not* producing food, there are no other relevant features of the act in question. But (as Singer notes, pp. 76–77), we can account for such a direct conflict pro and con an act by considering that such an act is too generally described. There are a number of kinds of acts of which (i) everyone's doing would yield some disutility, (ii) no one's doing would yield some disutility, but (iii) some limited number of performances would yield no disutility or perhaps (as in this case) some utility. What one is tempted to say is, the very fact that no one's producing food would have undesirable consequences is in effect a condition entailing that the *ceteris paribus* condition in the judgement against producing food is not satisfied, and vice versa.

One is tempted to say that whether or not other people are in fact producing food—and who, and how many—will determine which *ceteris paribus* condition is satisfied; but this leads to an eminently controversial topic, the relevance of others' behaviour, which I shall defer until much more ground has been covered.

It might also be supposed that the viciousness of the conflict between these two judgements turns upon their employing *one* description and in our applying one principle to this same description and therefore deriving, upon the *same* utilitarian grounds, judgements both pro and con an act in the same respect. This has been partly answered already. In the first place, there is not one, but rather two different descriptions in these two judgements. One is 'producing food' and the other is '*not* producing food'—what greater difference could there be? Similarly, the grounds are not the same in both cases, for in one case we rest the judgements upon the consequences of everyone's producing food, and in the other upon no one's producing (i.e. everyone's not producing) food. And there is clearly a difference between these two states of affairs. Such distinctions may be obscured if we fail to standardize the variant form of the generalization test. But if we assimilate the variant form back into the standard form, it becomes clear that two kinds of act are in question.

Further, one might suppose that entertaining two such closely related judgements, pro and con a certain kind of act, is at least pointless, since the two judgements obviously balance each other out. Here is where the restricted, non-comparative character of (G2) may mislead us. For, since (G2) is non-comparative, no consideration is given to relative disutilities, to differences in the undesirability of undesirable tendencies, to differences of degree in negative generalized utilities. And yet, one would expect that the tendency of 'not producing food' is far *worse* than the tendency of 'producing food'. Wouldn't things be in a much worse state if no one produced food than if everyone did? If so, then one judgement should not be regarded as balancing out the other—though, as I have said, (G2) does not clearly allow for such differences in the strength of judgements for or against an act. But then, this is a defect in (G2) which should not be legislatively obscured by reference to 'invertibility'. (Singer makes some reference to the two tendencies being 'equally disastrous',

p. 76; cf. p. 106. But, aside from the fact that they are undoubtedly not 'equally disastrous', it is a complete mystery where Singer finds a warrant for entertaining such comparative notions in applying his principle, and how, in general, they would be dealt with in his system. In any event, he deals only with cases where tendencies are equally undesirable and fails to allow that, since evils are more or less bad, the acts that produce them can in general be graded accordingly.)

Finally, let us observe that these conflicting weak judgements can be innocuously rendered in good reasons terminology:

Argument: Everyone has a good reason against producing food.
Counter-argument: Everyone has a good reason for producing food (i.e. against not producing food).

And now these appear harmless enough: merely a special case of reasons pro and con an act or a kind of action. There is nothing at all anomalous in such a conflict of reasons.

If there is no incompatibility in the 'producing food' case, there is even less chance of it in the case of someone lying in order to save another's life. For the conflicting judgements in the latter case are less similar:

Argument: If other things are equal, then it would be wrong for anyone to lie.
Counter-argument: If other things are equal, then it would be wrong for anyone not to lie-in-order-to-save-another's-life.

And this conflict can also be viewed:

Argument: Everyone has a good reason against lying.
Counter-argument: Everyone has a good reason to lie in order to save another's life.

We are now prepared to draw some general conclusions. Consider first the weak judgements that are derivable from a single principle: all such judgements are compatible, they do not conflict in any anomalous manner. This holds for general judgements, i.e. about kinds of acts, as well as for particular judgements, i.e. about particular acts.

Let us go further. It may have been observed that the sense I gave the *ceteris paribus* condition was limited to those features

of acts which are relevant for the application of the given principle (in these cases, a given form of utilitarian generalization). This is the restrictive interpretation of the *ceteris paribus* condition. There is also an inclusive interpretation, namely, one that requires for its satisfaction consideration of *all* relevant features of the act in question, whether relevant for the given principle or not. For example, there may be features of an act that are significant in respect of some other kind of valid moral principle, such as a principle of justice, but which are not significant for the application of any form of utilitarian generalization. On the restrictive interpretation, these features are not taken into account, but on the inclusive interpretation they are. Thus, on the restrictive, but not on the inclusive, interpretation of the *ceteris paribus* condition, the possibility exists of incompatible conflict between judgements derived from different forms of utilitarian generalization, or between judgements derived from some form of utilitarian generalization and from some other kind of principle. If we take the *ceteris paribus* condition inclusively, then no weak judgements are incompatible, whether derived from one or from more than one principle.

These considerations notwithstanding, it might still be thought possible that a single weak principle has incompatible implications in a special way. Suppose we satisfy the *ceteris paribus* condition of a particular weak judgement. This yields what I shall call the *resultant import* of a given weak principle with respect to a given act. The question now arises whether more than one such *ceteris paribus* condition can be satisfied, thus yielding conflicting (and therefore incompatible) strong judgements with respect to a single act.

We may conceive of the satisfaction of such a *ceteris paribus* condition at least as an ideal, namely, that all relevant features of the act are taken into account and are included in the description. (How much such a description must include depends, of course, upon how we interpret the *ceteris paribus* clause.) Ideally, then, the *ceteris paribus* condition of at least one particular weak judgement, from a given principle, regarding a given act, can be satisfied. But clearly, the *ceteris paribus* condition of no more than one such judgement can be satisfied. For the satisfiable clause is contained only in that unique judgement in which all

the relevant features of the act are mentioned. All other particular weak judgements concerning *that* act and derived from *that* principle will have incomplete descriptions and therefore unsatisfiable *ceteris paribus* conditions.

Suppose, for example, that there are six alternative courses of action in a given situation for a given agent: *a*, *b*, *c*, *d*, *e*, *f*. Then for each such act, e.g. *a*, it will be impossible to derive incompatible judgements from a single principle; for each act, one and only one strong judgement arises as the resultant import. This shows that the kind of inconsistency which accrues to strong principles on the method of rebuttals is spared weak principles.

What this fails to show, however, is that no incompatible judgements at all can be derived from a single weak principle. Consider now our earlier note of a special shortcoming of strong non-comparative principles, such as (G1). If (G1) is applied in a situation wherein all the alternatives, *a*,..., *f*, have undesirable tendencies, then incompatible judgements will result. This holds even if the descriptions are complete, even if—as we shall directly consider—the *ceteris paribus* conditions of corresponding weak judgements derived from (G2) are satisfied. Here we have several judgements condemning the several acts separately: '*a* is wrong', '*b* is wrong',..., '*f* is wrong'. And these are identical with the resultant import of (G2) with respect to the several acts. Each act is condemned; yet one thing is sure: that it cannot be the case that every alternative open is wrong.

But this particular anomaly, while not plausibly avoided in the case of strong principles, can be avoided for weak principles. For the anomaly arises from a restrictive interpretation of the *ceteris paribus* condition. In applying (G2), we are obliged to ignore the relative generalized utilities of *a*,..., *f* in any direct manner; they do not come into account in the derivation of particular weak judgements concerning *a*,..., *f*. But if, first, we interpret the *ceteris paribus* clause inclusively so that every relevant moral feature of the situation must be taken into account before the condition is satisfied; and if, secondly, we safely assume that not every alternative can be condemned; then the *ceteris paribus* condition of at least one of the weak judgements against *a*,..., *f* cannot be satisfied. That is to say, one of the moral conditions which must be satisfied for the derivation of strong judgements

in a given situation is simply that not every alternative act be condemned.

On this inclusive interpretation of the *ceteris paribus* condition, then, we should infer that all weak principles are compatible; that no incompatible judgements can be derived from weak principles of any kind, even when the *ceteris paribus* conditions of some such judgements can be satisfied.

Thus, the inclusive interpretation of the *ceteris paribus* condition may appear the reasonable one. And indeed, does not this interpretation best capture the real point of that condition? On the method of rebuttals, for the application of a *single* principle, the *ceteris paribus* condition allows for presumptions and rebuttals, based upon a certain model for the waging of moral argument. But I would be inclined to say that a method of rebuttals is inappropriate for the application of only one principle, and that it is rather more appropriate for the application of several different kinds of principles to a given act or set of alternatives.

We must distinguish two motives for incorporating a *ceteris paribus* condition in a given principle. In the first place, we may wish to preserve the compatibility of several different principles. Considerations of justice and utility, for example, may be accommodated in this way. Let us suppose that there are only two valid moral principles, one of justice and one of utility. If these are both *ceteris paribus* principles, then one can always be weighed against the other; the *ceteris paribus* condition would allow that weak judgements derived from one of these two principles could possibly be outweighed by judgements derived from the other. Or one of these principles could be regarded as primary, as always taking precedence; then that principle would be rendered strong and the other weak. But notice that, whenever either principle is applied, one must aim to take into account *all* the features of the situation which are relevant.

We might, secondly, want to incorporate a *ceteris paribus* condition into a given principle simply as a security against error. We might want to be able to apply the principle to a given act (as in the method of rebuttals) in respect of at least some but not necessarily all relevant features, in order to arrive at a tentatively acceptable conclusion which could be overridden if countervailing considerations of the same kind were later found

to be relevant. But surely our ultimate aim, in the application of any single principle, is to make our considerations as complete as possible, to approximate as closely as possible the ideal of completeness. Then all we require here is a notion of defeasible applications of a given principle, i.e that a given judgement for or against an act may possibly be replaced by a subsequent judgement which is based upon a more complete description of the act. We need not pretend that there is any special status owed to the original judgement; it need no longer be entertained; and the original, admittedly defective description need not be admitted as acceptable. We need not entertain any more than one judgement per act per principle, the one based upon the best, most complete description we have been able to achieve. For here we are concerned only with the resultant import of a given principle with respect to a given act. Thus, an ethical system which is composed of principles that supposedly cannot conflict (such as Singer's; see pp. 105–6) has no need for *ceteris paribus* conditions in its principles. In Singer's system, (G2) could easily be replaced by (G1)—provided his claims of consistency are sound.

What we require, then, is a method of applying the forms of utilitarian generalization which envisages a single, complete, and relevant description for a given act. Before examining such a method, however, let us note two other difficulties accruing to the method of rebuttals.

First, the condition of relevance is too weak. There are descriptions which satisfy it and yet should not be regarded as relevant for the application of any form of utilitarian generalization. I shall discuss this more fully later. Suffice it to say for the present that my objection rests upon a distinction which is obscured by the condition of relevance in the method of rebuttals. That condition requires that the properties of acts which are mentioned merely be associated or correlated with significant generalized utilities. But the condition should require that the properties mentioned are those in virtue of which generalized utilities obtain.

Secondly, the method of rebuttals can yield extremely misleading, if not perverse, results. Consider, once more, a case of lying in order to save another's life. Let us assume that no other features of the particular act are relevant, and also that (i) the tendency of 'lying', but not of its complement, is undesirable, and

(ii) the tendency of 'lying in order to save another's life' is desirable, and that of its complement is not undesirable. Now it seems clear that we should regard the more complete of these two descriptions as the more adequate one, and that—on any form of utilitarian generalization—the particular act should not be condemned. But on the method of rebuttals we are obliged to ignore the more complete description, we are obliged to disregard the special circumstances (or motive) surrounding (or underlying) the lie. We can derive a judgement against the act, but none for it; the presumption against the act cannot be rebutted. And this presumption against the act would be conclusive for the application of (G1) and would be the resultant import for the application of (G2).

What has gone wrong is that we have allowed judgements based upon partial descriptions when these are admittedly incomplete. The way out is to allow consideration only of the complete (or, practically, most complete) description and to entertain only those judgements which are derivable on the basis of such descriptions. And in a case such as the foregoing, we should conclude that no judgement against the act could properly be derived.

But I have anticipated later developments, for we have not yet discovered how a single description can, in principle, be determined—and most important, we have not yet discovered an adequate test for relevance. To these problems we now turn.

c. *General Utilitarian Properties*

We require a method for assessing descriptions of acts for the application of utilitarian generalization, a method for determining whether each element in a given description is relevant and for determining whether any relevant predicates have been omitted. The criterion should, moreover, be quite general, first, in that it should acknowledge differences in degree as well as differences in the quality of generalized utilities. And secondly, the criterion should allow us to consider all features of an act that may be considered relevant for the application of any form of utilitarian generalization, so that it does not mislead in the manner of the method of rebuttals just noted. We should be able to determine the proper description of an act based on the kind

of properties of acts that are relevant for this kind of principle. Then a form of utilitarian generalization could be applied to an act in respect of that description, yielding a judgement if the antecedent condition of that form were satisfied.

It will be helpful now to consider one criterion which actually has been proposed for use with a form of utilitarian generalization—one that does not acknowledge a plurality of descriptions for a given act. Jonathan Harrison has proposed such a criterion (in 'Utilitarianism, Universalisation, and Our Duty to Be Just'); he does not fully formulate it, however, so we shall have to get at it through his own examples and comments. The main example which Harrison considers is this:

> Suppose, for example, that a red-headed man with one eye, a wart on his right cheek, and a mermaid tattooed on his left fore-arm, were to tell a lie on a Tuesday. It might be argued that it was quite permissible for him to have told this lie, because his action in telling the lie belongs to the class of actions performed on a Tuesday, and the consequences of the general performance of actions on a Tuesday is indifferent. (Harrison, p. 114.)

But such a claim would be unwarranted, Harrison urges, for we would not then be considering the proper sort of action. The action could be further relevantly specified:

> For the class of actions performed on a Tuesday contains within itself a number of sub-classes: deceitful actions on a Tuesday, self-sacrificing actions on a Tuesday, revengeful actions on a Tuesday, and so on. The consequences of the general performance of actions in these sub-classes will differ both from one another and from the consequences of the general performance of that wider class which is the genus. (p. 115.)

Apparently, the test for the possibility of further relevantly specifying an act is whether there are any species of the given description the tendencies of which differ in value from one another or from the tendency of the genus-description. If there are such species, the description of the given act must include the species under that genus to which the action belongs.

There is also a test to determine the relevance of any given (initial) description (or part thereof). For some of the properties mentioned may not really be relevant:

> The class of actions, 'lies told by one-eyed, red-headed men, with warts on their right cheeks, and mermaids tattooed on their left fore-arms', is a wrong one because it can be 'irrelevantly generalized'; that is to say, by subtracting characteristics such as 'being an action performed by a one-eyed man' I can obtain more general classes of actions, the consequences of the general performance of which do not differ from the consequences of the general performance of actions belonging to it. (p. 116.)

This test appears to be the reverse of the one for further relevant specification. A given description (or part thereof) can be irrelevantly generalized if, taking that description as a species, there are no differences in value between the tendency of that species and (i) the tendencies of the other species of the genus, (ii) the tendency of the genus itself.

It is interesting to note, incidentally, that, whereas Harrison is inclined towards fixing a single determinate acceptable description for each act on such a method, Singer is not so inclined, and accepts a plurality of acceptable descriptions. The problem of inconsistency may suggest why this is so. Harrison's principle is the analogue of Act-Utilitarianism, it is a strong, positive, comparative form of utilitarian generalization (with the 'If, and only if' introduction), which in my reconstruction of it goes:

(GU) If, and only if, the consequences of everyone's doing a certain sort of thing would be worse than those of some alternative, then it would be wrong for anyone to do such a thing.

Since (GU) is strong, we would not want to allow a plurality of acceptable descriptions, lest the spectre of inconsistency rear its head once more. Moreover, since (GU) is comparative, its application would be somewhat fantastic unless fixed descriptions were assigned each act in a set of alternatives. But as we have noted, on the other hand, Singer's (G2) is a weak principle; and such a principle, exempt as it is from the problem of

inconsistency, may unfortunately incline one towards a plurality of descriptions.

Now it is clear that Harrison intends (or should intend) that his criterion be complete: it must provide a method for testing for the relevance or irrelevance of any and every description that is true of a given act. Unless this provision is made, one could not hope to determine a single certified description for each act. But then the criterion needs more careful formulation.

Let me suggest the kind of procedure that, if it were possible to follow, would provide for completeness and generality. First, all properties of a given act, all descriptions that are true of it, would have to be listed; or at least some provision must be made that they be exhaustively considered. This is essential, for the criterion is to be the basis for determining the relevance of descriptions; if some were not to be considered, our final description might not be complete. Such a listing would, of course, be incompletable, for there are innumerable properties of each particular, concrete action. The second step, moreover, would be even more formidable: we would have to consider combinations of properties—not only combinations composed of the properties of the act itself, but all combinations that could be developed from them, taking each as a genus-description. For we are to determine, on this method, whether there is any difference between the value of the tendency of any genus-description and the value of the tendency of any species under that genus. It should be observed that, if there were only n properties of the given act, there would be 2^n-1 combinations of those properties alone. But, in the first place, the number we begin with is unlimited; and secondly, we are not restricted merely to those combinations. Thus, the number of tests would be far greater than 2^n-1, where n is not as small as any positive integer.

The fact that such a test-procedure could not actually be carried out to completion should not be surprising. The complexity of such a task is a consequence of the complexity of human actions, and its difficulty is also a function of our limitations as finite beings. But any plausible moral theory—especially if it includes teleological elements—would have to face difficulties of this kind. The fact that these difficulties are reflected in our development of Harrison's criterion is less

a mark against it than evidence of realism in Harrison's approach.

As I have already suggested, it would seem plausible to regard the completed application of such a criterion as an ideal which cannot practically be achieved; although, in particular cases we can be quite certain and justified in supposing that all features of acts which make a great deal of difference to our results have been taken into account. Accordingly, any given description based upon an actual application of this criterion (to the best of our abilities) and any judgement actually derived on the basis of such a description may be regarded as defeasible, held tentatively as correct though subject to the contingency that it might be replaced by a more adequate description or by a judgement derived upon the basis of the more adequate description.

The main problem regarding this criterion is not its complexity, nor the incompletability of its application, but rather the fact that it is mechanical. What I mean is this: both the method of rebuttals and this criterion (developed out of Harrison's suggestions) fail in the same way; they fail to incorporate a material condition of relevance, a condition relating to the *nature* of properties which are relevant for utilitarian generalization.

Let us suppose that, as in Harrison's example, a red-headed man with one eye, a wart on his right cheek, and a mermaid tattooed on his left forearm were to tell a lie on a Tuesday. How should such an act be viewed? Can it be at all relevant, for example, that the man has red hair, and that his act is performed on a Tuesday? Let us suppose, further, that in applying Harrison's criterion we were to find the tendencies of actions that included 'performed on a Tuesday' as a condition to be consistently worse (or better) in value than the tendencies of actions that were similar on all other counts but not performed on Tuesdays. This is a possibility worth considering, for Harrison does not demonstrate that such a finding could not result from the application of his tests. He merely assumes that it would not make a difference to the value of the tendency whether 'performed on a Tuesday' were or were not included. But if our tests showed that adding or subtracting it made a difference in the manner required, we should infer from Harrison's criterion that 'performed on a Tuesday' is a relevant description of an action.

Such results would be extremely disquieting and would throw doubt upon the adequacy of the criterion. We would at least doubt the validity of the tests made and would seek some explanation. For we do not regard 'performed on a Tuesday' as a legitimate candidate for inclusion in the relevant specification of an action from a moral point of view. More important than our preconceptions of moral relevance, however, is the fact that 'performed on a Tuesday' could not serve as part of the relevant description of an action for the application of a form of utilitarian generalization. Why is this so?

Pure teleological principles are concerned, ultimately, only with the values of the consequences of actions. Thus, when pure teleological principles are applied, particular actions may be viewed only with respect to their teleologically significant properties, that is, those properties *in virtue of which* actions produce utilities or disutilities. The forms of utilitarian generalization are pure teleological principles. Therefore, the only legitimate candidates for inclusion in the description of an action for the application of a form of utilitarian generalization are causal or consequentially significant properties, that is, those properties in virtue of which actions have effects. Thus, *general utilitarian properties*—those properties that are relevant for the application of utilitarian generalization—are causal properties in virtue of which the universal performance of acts of that kind would produce some utility or disutility.

The description 'performed on a Tuesday' is not a legitimate candidate for consideration as a general utilitarian description simply because we cannot, consistently with our other beliefs, hold that it is the sort of property in virtue of which actions have effects, no less good or bad effects. The temporal location of an action, as such, is not causally relevant. The time at which an act is performed may indeed make a great deal of difference to the value of its effects, but not in virtue of the time as such. It might very well happen that the effects of acts performed on Tuesdays are consistently worse (or better) than those of similar acts performed on other days. But if so, the other acts could not be similar in all other relevant respects. If they seem so, then we must infer that there is an obscured, an as yet undiscovered causal connexion between some feature of the situation on Tuesdays and the value produced.

In other words, there is an important material condition for the selection of predicates as relevant for the application of a form of utilitarian generalization. It is not enough for there to be mere correlations between changes in a description and differences in the values of the tendencies; there must also be a causal connexion between the property mentioned and those effects accounting for the gain or loss of utility.

If this condition is neglected, *ir*relevant specifications can be certified as relevant. For example, if everyone with red hair, only one eye, a wart on his right cheek, and a mermaid tattooed on his left fore-arm were to tell lies, less undesirable consequences would result than if everyone (without further qualification) were to tell lies. And if everyone were to tell lies on a Tuesday, less undesirable consequences would result than if everyone were to tell lies. The reason less undesirable consequences would result is simply that fewer lies would be told; the added specifications in these cases restrict the class of acts by restricting the number of individuals who would have occasion to perform them or the circumstances in which they might be performed. But the differences in utility (or disutility) involved are not causally or teleologically related to temporal location as such and are most likely not so related to the colour of one's hair, the number of good eyes one has, and so on. The *causal* connexions involved do not bear upon these properties as such; these properties have no connexion with the values of the effects in question. The total value of the effects varies simply because more or fewer instances of the sort of action in question (lying) would be performed. It makes no difference, that is, what the properties happen to be in such cases. Any number of arbitrary descriptions could be used to the same effect.

This is, perhaps, what Singer is groping towards with his minor condition against 'reiterability' (pp. 80–83)—that is, that it simply makes no utilitarian difference which egocentric particulars or indicator-words are mentioned in a description, and that the results can be misleading when such properties are taken into account, because there is no connexion between the utilities in question and these properties as such. But Singer simply fails to account for the material irrelevance of such properties and thus he fails to explain or properly deal

with the notions of arbitrariness and inessentiality of specifications.

The material irrelevance of such properties is based upon causal considerations. The mistake of neglecting this condition would be to infer the existence of causal connexions from mere correlations. The fallacy involved, implicitly, when such inessential or irrelevant properties are regarded as relevant, is *non causa pro causa*. Unless we take into account causal connexions we would be led to infer the relevance—the causal significance, therefore—of 'performed on a Tuesday', 'having red hair', 'having thirteen toes', and so on.

The notions of a causal connexion and of causally significant properties should not be understood in this context in any special, restricted sense. We are not concerned here merely with brute physical characteristics and effects. If performing an act from a certain motive has the effect of producing some utility or disutility, then the property 'performed from such and such a motive' is to be taken as teleologically and, therefore, in this broad sense, as causally significant. Similarly, at least some of the value produced may be the quality of psychical experience. To the extent that such qualities are the *effects* of certain properties of actions, those properties are to be regarded as having teleological significance.

In view of this broad use of the term 'causal', it might appear that we should admit an exception to the foregoing exclusion of spatial and temporal predicates from the relevant description of acts for the application of utilitarian generalization. Suppose, for example, that some persons reacted strongly to an act's being performed on a Tuesday in virtue of, say, great value or disvalue placed upon acts for that reason, i.e. that they are performed on a Tuesday. If these people react to a belief that an act is performed on a Tuesday, then there is some ground for claiming that the fact that the act is performed on a Tuesday in conjunction with certain attitudes or dispositions of theirs, is psycho-causally connected with some utility or disutility—perhaps displeasure or great joy. If they merely react to the performance of the act itself in, and in virtue of, its temporal location, then we should be even more strongly inclined to include 'performed on a Tuesday' in a complex description which can go into the relevant description

of some such acts. But these cases are, strictly speaking, no exceptions to the foregoing argument. For the main point urged above is simply that we must consider only those determinate properties of acts in virtue of which utilities or disutilities are produced. And the court of appeal will be the facts of the case, the chain of events leading to the utilities and disutilities actually produced. And while spatial and temporal predicates may conceivably enter into descriptions through the back door, as it were, they will enter in as a logical consequence of causal, socio- or psycho-causal, laws which themselves are not spatially or temporally restricted.

I have spoken of Harrison's criterion and the method of rebuttals as mechanical. They are mechanical in the sense that the tests to be followed are based upon rules which give no cognizance to causal connexions. They may also be regarded as mechanical in the further sense that both Singer and Harrison seem to imply a practical or theoretical limit to the number of steps in the process of testing for relevance. Our comments upon Harrison's criterion cast some doubt on this possibility. But in any event, such a mechanical procedure cannot be relied upon to distinguish between correlations and causal connexions—in effect, between apparent (but illusory) and real general utilitarian properties.

There is, therefore, no simple formula for determining relevance for the application of a form of utilitarian generalization. Reference must be made to our scientific knowledge and view of the world. Moreover, the class of general utilitarian properties will be determined, not only by the content of this body of knowledge and theory, but also by value-criteria, that is, the criteria which can be applied in determining the value of states of affairs. The determination of relevance is, therefore, in principle a highly complex matter.

Thus, the legitimate candidates for inclusion in the description of an act for the application of any form of utilitarian generalization are causal properties only. General utilitarian properties are a sub-set of these (though not necessarily a proper sub-set), relative to a set of value-criteria applied to the tendencies of actions. Descriptions should include none other than general utilitarian properties.

As I have already urged, it is not enough merely to take some

general utilitarian properties of an act into account. Ideally, we should take all such properties into account: certainly, if we are settling upon one description per act. To the earlier considerations to this effect we may add this: failing to make a description complete for the application of a form of utilitarian generalization is akin to failing to take into account all the utilities and disutilities attributable to a given act when a form of simple utilitarianism is applied.

To summarize: our approach is to decide which similarities and dissimilarities among acts should be taken into account by reference to the nature or content of the substantive principles being applied. An account can be given of the kinds of similarities and dissimilarities which should be taken as relevant for the application of a form of utilitarian generalization. There is, in principle, no difficulty therefore in applying a form of utilitarian generalization; although there is here, as with so many kinds of principles, much room for error in practice.

III

EXTENSIONAL EQUIVALENCE

WHY should we suppose that it makes an important difference how the principle of utility is applied? Why should philosophers be inclined to assume that acts are assessed differently when they are assessed in respect of their generalized instead of their simple utilities? These are the questions we shall seek to answer now. As I have already suggested, the issue may be formulated as that of extensional equivalence. The non-equivalence thesis is that there are some cases in which analogous forms of these two kinds of utilitarianism do not yield substantively identical or equivalent judgements.

But first, how does one argue for or against a thesis like this? It might be attempted by examining specific cases, but this is a line I shall not adopt. Notice that an argument for non-equivalence requires only one case in which two analogous principles differ. But in general, an issue like this cannot reliably be argued from examples. For one thing, they depend upon extremely complex factual and evaluative considerations. Secondly, the proponents of utilitarian generalization do not want merely *some* non-equivalence, but rather differences accommodating the criticisms against simple utilitarianism. Thus, the examples chosen must be not only rigorously developed but also typical and crucial. But for the sake of a conclusive argument I cannot rest my case upon examples, for I shall argue for equivalence, and no limited set of convergent examples can constitute a proof of that thesis.

I shall therefore hazard an argument on *a priori* grounds. I shall first examine the line of reasoning which would lead one quite naturally to suppose that non-equivalence obtains. This line of reasoning is suggested by the two writers who have been most explicit in arguing for non-equivalence, R. F. Harrod and J. Harrison. The considerations these writers adduce seem to go to the heart of the matter. But the line of reasoning to be developed here should not, strictly speaking, be attributed to Harrod or

Harrison. In the first place, their arguments are not explicitly complete, and I do not want to claim that either writer would necessarily fill in the gaps exactly as I have done. Secondly, their arguments are meant only to replace Act-Utilitarianism with its general utilitarian counterpart, and thus they are not explicitly general. And finally, their contrasting principles seem to be probabilistic, a fact which suggests complications and further arguments which we shall not examine in detail, though I shall allude to some.

The line of reasoning which leads towards non-equivalence rests upon a causal factor, a factor which I shall characterize and the normative significance of which seems to have been misunderstood by Harrod and Harrison. This factor is what I shall call the *threshold* phenomenon. The crucial consideration in its analysis is the social context, the general pattern of behaviour, in which an act is performed. From the agent's point of view and for the purpose of describing a given act, this factor concerns to what extent others are doing *the same*. Now it happens that a number of writers have had reservations about the relevance of such a factor for the application of utilitarian generalization (or for the application of rule-utilitarianism). It may seem *wrong* to take such circumstances into account, or else paradoxical or contradictory. I shall contend that such circumstances can and must be considered, that they must be included in a complete general utilitarian description of a given act. And when one reckons them in, one secures the argument for equivalence.

A. *Linearity and Thresholds*

What conditions would be necessary for the truth of the non-equivalence thesis? The primary difference between the two kinds of principles is the manner in which the test of utility is employed when we apply them. That is, they require us to assess acts as functions of their simple or generalized utilities, respectively. It would seem, therefore, that non-equivalence should turn upon some irregularity in the relations between these two standards. Harrison, for example, claims that there is a *qualitative difference* between the simple and the generalized utilities of some acts:

> There are some actions which we think we have a duty to do, although they themselves produce no good consequences, because such actions would produce good consequences if they were generally practised. There are some actions which we think we have a duty to refrain from doing, even though they themselves produce no harmful consequences, because such actions would produce harmful consequences if the performance of them became the general rule. (p. 107.)

If this were in fact the case then there surely would be non-equivalence, for when such a condition obtained there would be a divergence in the implications of analogous non-comparative principles. Such principles require us to assess acts on the basis of their absolute utilities (i.e. without regard for the utilities of alternatives) and on the basis of their quality only. Thus, if there is such a qualitative difference between the simple and the generalized utilities of but one act, there is non-equivalence.

But again, why should we suppose that such is the case? It will be helpful to exploit some of Harrod's and Harrison's examples, since these are typical of the kinds of cases with which philosophers who have employed generalization principles have been concerned. Harrison argues:

> I think I have a duty to vote for that person whose party I think would govern the nation best, although I do not think that the addition of my vote to the total number of votes which are cast for him is going to make any difference to the result of the election, simply because I realise that, if all his other supporters were to do as I do, and fail to go to the polls, the man would not be elected. (p. 107.)

We shall disregard some features of this for the present—for example, the significance of the belief-qualifications 'I think . . .' and 'I do not think . . .'. Let us suppose simply that what is claimed is a qualitative difference between the *actual* simple and generalized utilities of a given act. Harrison's voting or not voting is indifferent in so far as it would make no difference whether he voted or not; thus his own act is claimed not to have undesirable effects. But if everyone did the same, if everyone who would want Harrison's man in, failed to cast a favourable ballot (assuming Harrison's man is the best candidate, has the best

party, &c.), then bad consequences would result; thus the the tendency of (roughly) failing to vote for that person whose party [one thinks] would govern the nation best is undesirable, whereas at least one instance of that kind has an indifferent simple utility.

If such a condition of qualitative difference obtains, then it follows that a more general condition also obtains—what I shall call utilitarian *non-linearity*. Non-linearity is a possible characteristic of the relation between the generalized and the simple utility of a given act. It is the characteristic of such a relation when the generalized utility is not a *linear function* (in a sense now to be explained) of the simple utility. Let G represent the generalized utility of a given act—the value of everyone's doing the same in respect of a given description; and let S represent the simple utility. There will be n occasions for doing the kind of act in question; that is, if everyone who had occasion to do so on each such occasion did the sort of act described, that sort of act would be performed n times. Linearity then obtains when $G = n \times S$; any other relation is non-linear.

Thus, non-linearity is a necessary condition of a qualitative difference between G and S; for if G and S have different qualities, then they cannot in any way be proportional, and hence not proportional in this special way. Consequently, if non-linearity is precluded, so is such a qualitative difference, and so, perhaps, is the argument for non-equivalence. But there may be a more direct and intimate relation than this between non-linearity and non-equivalence. Harrod supposed that there was, and I shall indicate his mode of argument now.

Harrod claimed first (p. 147) that there are 'hard and fast obligations' which simply cannot be accounted for by Act-Utilitarianism: obligations to keep promises, to tell the truth, &c., which are effective or binding even when such courses of action yield worse effects than their alternatives. Now Harrod clearly does not mean that these are absolute or unexceptionable obligations, as if one must always tell the truth and keep his promises regardless of the circumstances. He means rather something like prima facie obligations—but not merely any sort of prima facie obligations. It is not just that one normally has a good reason to tell the truth or to keep one's promises. These obligations

are supposed by Harrod to be *stronger* than they appear to be on Act-Utilitarian grounds. In other words, as I shall show, (GU) provides stronger reasons than (AU) for doing such things.

But this argument must be reconstructed from Harrod's presentation. The argument develops somewhat along the following lines. Harrod claims that the total, cumulative effects of the performance of acts of some kinds are enhanced in magnitude when such acts are generally performed and hence are not proportional to the effects of particular such acts:

> *There are certain acts which when performed on n similar occasions have consequences more than n times as great as those resulting from one performance.* And it is in this class of cases that obligations arise. It is in this class of cases that generalizing the act yields a different balance of advantage from the sum of the balances of advantage issuing from each individual act. For example, it may well happen that the loss of confidence due to a million lies uttered within certain limits of time and space is much more than a million times as great as the loss due to any one in particular. Consequently, even if on each and every occasion taken separately it can be shown that there is a gain of advantage (the avoidance of direct pain, let us say, exceeding the disadvantages due to the consequential loss of confidence), yet in the sum of all cases the disadvantage due to the aggregate loss of confidence might be far greater than the sum of pain caused by truth-telling. (p. 148; Harrod's italics.)

For the loss of confidence in communication is intensified by the very frequency of lying.

This passage is cited at length because it is clearly non-probabilistic and thus supports our rendering of the non-equivalence argument. None the less, although Harrod had little else to say on the matter, this clearly does not involve an explicitly complete argument for non-equivalence. In the following I shall suggest one way of extending the argument starting from an argument for causal non-linearity; later I shall deal with an alternative approach which starts from utilitarian non-linearity directly.

It might be claimed simply as a fact that there is causal

non-linearity for some acts. Let **T** represent the tendency of an act, i.e. the total consequences resulting from n performances of acts of a certain kind of which the act in question is an instance; and let **E** represent the effects only of that act. Causal non-linearity would obtain if, for example, **T** were greater than $n \times \mathbf{E}$. If there is causal non-linearity, then it is reasonable to assume that there is utilitarian non-linearity. In the present case, for example, assuming that utility is simply correlatable with or proportional to effects, **G** would be greater than $n \times \mathbf{S}$. Now we may combine this condition with Harrod's claim about 'hard and fast obligations', i.e. that they correctly appear to be stronger on the general than on the simple utilitarian account. We may take this as a claim that the revised principle, (GU), provides stronger reasons for or against acts of the kinds in question. Then, if **G** is sometimes greater than $n \times \mathbf{S}$, the generalized utilities are sometimes disproportionately greater than the simple utilities for these cases. Whereas, ordinarily the process of generalizing would merely provide what Broad has called a 'moral microscope' which can serve to sharpen our awareness of the results of acts (Broad, pp. 382–4); on the other hand when **G** is greater than $n \times \mathbf{S}$ and we assess these acts according to **G** instead of **S**, the enlargement as it were is greater. Under such conditions, if (i) **G** is positive (a desirable tendency) or (ii) **G** is negative (an undesirable tendency), then general utilitarian reckoning would seem to yield proportionately stronger reasons than simple utilitarian reckoning does (i) for such acts or (ii) against them, *vis-à-vis* other acts for which there is no non-linearity. And if there are such stronger reasons pro or con acts of certain kinds, then one's obligation to perform or to refrain from performing them would be so much stronger, it would be harder to rebut the presumption for or against them, as the case may be. Thus such acts would most likely be condemned or prescribed, as the case may be, more often, or under more circumstances, by general utilitarian than by simple utilitarian considerations.

There are still gaps in this argument. One must show, for example, that non-linearity obtains only for those kinds of acts which one has reason to believe are 'hard and fast obligations'; moreover, that it obtains in the requisite directions and to the requisite degree. But I shall not pursue the matter further into

these details. For the important and interesting problem is non-linearity itself: *Is* there such a condition? *Can* there be such a condition? If so, under what circumstances? I shall go on, after a brief digression, to examine these questions.

But first I wish to make clear once more what is meant by 'non-equivalence' in order to forestall a kind of argument which might obtrude but which is after all just beside the point. A proponent of revisionistic utilitarianism might be tempted to argue for 'non-equivalence' in this way. The results of everyone's deliberating and acting upon, that is, of everyone's attempting to follow, Act-Utilitarianism would be different from and worse than the results of everyone's attempting to follow the analogous form of utilitarian generalization. For example, if we fail to generalize we overlook certain effects that could only be produced by a general practice of similar acts; and thus avoidable harm could be produced were everyone to attempt to follow Act-Utilitarianism. It might be argued that reckoning from case to case is hazardous, and that a general attempt to follow a set of rules that are formulated in an awareness of the contingencies of general practices would have better effects on the whole than general observance of Act-Utilitarianism. (See below, Chapter IV.)

An Act-Utilitarian, on the other hand, might claim that reckoning on general utilitarian grounds would prevent one from making the best of circumstances as they are (without falsely supposing that others are or will be doing the same); and in particular that the argument for rules on utilitarian grounds is misguided and amounts to 'rule-worship'. (See below, Chapter V.)

But it should be emphasized that we are not here concerned with what might go wrong in the application of one or the other kind of utilitarian principles. We are not concerned with how people might overlook certain kinds of effects and thus mistakenly infer certain judgements from their principles; nor with how much safer it might or might not be to apply a simple set of simple rules instead of reckoning from case to case. Such considerations do not add up to arguments for non-equivalence in the sense I have been giving that term. For we are only concerned with the actual import of these various principles—as opposed to what we might mistakenly infer from them; we are concerned with

what they actually imply—as opposed to the judgements we might be inclined to attribute to them, especially under conditions of imperfect theoretical understanding, lack of information, strong temptation, and so on. I believe that this distinction can be made, that it is important, that it is essential to an understanding of the significance of various kinds of utilitarianism.

To resume. Let us note that there is something paradoxical about a condition of non-linearity. Consider causal non-linearity, wherein T does not equal $n \times E$. Under such a condition it would seem that some effects arise uncaused or appear mysteriously. It would seem, when T is greater than $n \times E$, for example, that effects which are not attributable to any of the n acts (nor of course to anything else) arise when n such acts are performed. These 'uncaused' effects amount to the difference between T and $n \times E$. If T were less than $n \times E$, on the other hand, we should have a similar quantity of effects produced by the n acts which are not a part of T and thus somehow disappear—and disappear, not in the sense in which negligible effects can be dissipated, but in some causally objectionable sense.

This apparent paradoxicality will be dispelled by a more thorough causal analysis of certain features of general practices. Harrod's argument suggests that certain kinds of effects which can be produced by the general practice of some acts are simply not producible otherwise. It is these special, or extra, effects into which we shall now inquire.

Let us exploit another of Harrison's examples in developing our causal analysis:

> I refrain from walking on the grass of a well-kept park lawn, not because I think that my walking on the grass is going to damage the lawn to such an extent as to detract from anybody's pleasure in contemplating it, but because I realize that, if everybody else who walked in the park were to do likewise, the grass in the park would be spoilt. (p. 107.)

Here again we shall disregard the implicit probabilistic qualifications that are suggested by the expression 'I think that . . .'. We are concerned, not with Harrison's expectations about the effects of his act or of a general practice, nor with their probability, but rather with the question whether this act, or the general practice

of such acts, will in fact damage the grass. We are concerned with the utilities of the acts, not with their probable or supposed or expected utilities.

Now whether any damage—or damage sufficient to detract from someone's pleasure—is done to the grass, is a function partly of the number of crossings, partly of the time-interval involved, and partly of certain background or standing conditions which for the sake of argument we can assume to be quite stable, such as climatic and soil conditions, the degree and kind of maintenance, and so on. If we assume the stability of these background conditions, then we may say that *frequency* (as opposed to the mere number of crossings) is a critical factor. A given lawn may be crossed a certain number of times within a given time interval without appreciable damage resulting, that is, without damage appreciable enough to detract from anybody's pleasure in contemplating the lawn. Thus, since appreciable damage does not result from one such act, it is prima facie plausible to claim that the effects of such acts are negligible and most important not undesirable; S is not negative. But if, under the same background conditions, and within the same time-interval in which one or a few conjointly harmless acts could be performed, very many were to cross the lawn, then the grass would be (appreciably) damaged, and less pleasure—if any at all—would be derivable from its contemplation. It therefore also seems plausible to regard the tendency of such an act as undesirable; G is negative.

(For simplicity's sake I shall deal with this case as if the only critical causal and evaluative factors involved relate directly to lawn damage; I shall later show how such a case can be viewed in another way. It should also be observed that reference to frequency is not always sufficient; this will be remedied as the causal analysis is developed. The present approach will hopefully shed some light upon unclear facets of the arguments in the literature.)

One interesting thing about many such cases is that a *universal* practice is not required to produce the special effects. Some of those who would have occasion to cross the lawn may fail to do so, and this is compatible with the lawn's being appreciably damaged. What is required is that the practice be general beyond

a certain point. It may take the efforts of two oarsmen to keep a boat moving—if fewer than two pull, their craft will be at the mercy of currents and tides. But if there are more than two available, some can abstain without affecting the result in that respect. On the other hand, if there are only two oarsmen in the boat, then (neglecting the different weight factors) everyone (who is in a position to do so) must pull. Harrison thought that these two cases represented significantly different causal conditions. It is clear, however, that contingent circumstances related only to the number of occasions on which an act can be performed divide these types of cases. But the case in which there is a surplus of man-power, as it were, is probably the more general, and we shall examine this type of case in greatest detail.

Thus, ordinarily a few individuals can cross the lawn without doing appreciable damage; a few supporters can abstain without affecting their candidate's chances; a few lies can be told, a few promises broken, a few goods sold on the black market; a few citizens can evade their taxes and military service or exploit rationing restrictions—to take but a few of the commonest examples of the application of utilitarian generalization—without undesirable effects resulting on the whole, provided that most others do not do such things. Indeed, it may be plausible to claim that in some cases better results on the whole would be produced if some such exceptions occurred. But the generalizers want to argue that, because it is not the case that everyone (or even most people) can do such things without undesirable results, no one should do them (except perhaps in extraordinary circumstances). By tying the rightness and wrongness of acts to generalized instead of simple utilities, and by supposing the irregular G/S relation, a utilitarian argument against such acts is offered.

Another important feature of such practices as have special utility or disutility is that these results do not depend upon the performance of any particular sub-class of acts from among those which could be performed. Given the requisite causal background, the extra effects depend solely upon such factors as frequency. For example, Jones passes the lawn regularly and thus he might or might not cross it; perhaps it would save him steps to do so. But whether he happens to cross it can be immaterial in the present context. For the lawn can be damaged

when he refrains, if others cross it frequently enough; and it can remain undamaged when he crosses, if few others do the same. But on the other hand it might be undamaged when he refrains and damaged when he crosses—and he might contribute to that damage. Moreover, it is also possible for an individual's act in some circumstances (though perhaps not in a lawn-crossing case) to be critical, to determine whether or not the extra effects will be produced.

Given the requisite background conditions, then, the crucial factor determining whether such extra effects (and value) are produced is the pattern of performance of acts of the kind in question, i.e. whether or not the over-all performance is *dense* enough to complete the causal requisites of the extra effects. ('Density' is a more general term than 'frequency', designed to apply as well in non-temporal cases, such as voting, which we shall examine.) The qualitative change from the negligible to the appreciable, and in general the accumulation, expansion, and enhancement of effects depend upon this factor, that is, upon *how general* the practice is. From the point of view of the agent and for the purpose of describing a particular act, this crucial factor is to what extent others are or will be doing *the same*: the pattern of others' behaviour.

This leads us to a general characterization of the causal phenomenon. For many kinds of act there are *thresholds* which must be passed before effects of certain kinds or of certain magnitudes can be produced. For many kinds of acts, in other words, the curves describing the relation between total effect and the number of performances are irregular or at least not straight lines. (I am neglecting the step-wise character of curves for temporally-ordered or sequential acts.) In the lawn-crossing case, for example, the threshold may be indicated roughly as that point at which, if crossing became more frequent, the lawn would be damaged. This damage (which is no longer negligible) may be called a *threshold effect*. In such a case the threshold effect itself has perhaps no intrinsic value but is instrumental to a loss of pleasure (which we are supposing for the sake of argument to be a disutility). In the case of voting, the threshold is the number or proportion of votes required to elect a certain candidate or to pass a given bill; and the threshold effect in question is the

result—election or defeat, passage or loss—which again has primarily instrumental value. In Harrod's example of lying, it would seem that 'certain limits of time and space' are necessary conditions of the disproportionate loss of confidence in communication (which respectively constitute the threshold and the threshold effect).

Clearly, this phenomenon is basically causal in character; it is a matter of the varying efficacy of acts under varying conditions. It will be helpful to continue to distinguish, even more carefully, the causal and the normative issues. Harrod and Harrison, who come closest to identifying the causal factors, simply fail to separate the issues. We shall seek an answer to the normative questions through the avenue of *a priori* causal analysis.

We may set the stage by noting that the threshold phenomenon is not at all uncommon in human affairs. This indeed enlarges its potential normative significance. Many kinds of acts have associated thresholds in the sense that, if enough such acts are performed, the effects of the several acts will be modified appreciably. These acts, their thresholds, and threshold effects can be classified in many ways. We shall find it particularly useful to distinguish between sequential and non-sequential acts—a contrast exemplified in the two kinds of voting that we shall examine, namely, voting by roll-call and simultaneous balloting.

If the threshold phenomenon is not uncommon in human affairs its analogues are commonplace in nature; a significant aspect of natural science consists of their examination. We might even take as a natural model for the threshold phenomenon the change of state of physical bodies, such as water boiling after regularly rising in temperature to a certain point while a constant amount of thermal energy is transferred.

Nor are threshold effects (or something akin to them) confined to the general practice of acts that are, in a straightforward sense, similar. It also takes a number of people doing *different* things to produce certain kinds of effects. On a team, in an orchestra, in a work-gang, on an assembly-line, various individuals, performing different but related tasks, contribute to the production of effects that could not be produced otherwise. A natural model for such phenomena is the combination of chemical elements into compounds that have characteristics qualitatively different from

those of the elements joined. There is nothing mysterious or unusual about the kinds of effects noted by Harrod.

Now it may be difficult in theory or in practice to determine the nature of a given threshold—or perhaps even whether there is one associated with a given kind of act. Moreover, a threshold need not be a point; there may be a critical range of acts (instead of just one act) which must be passed before threshold effects can be produced. Furthermore, the actual threshold-relations may be quite complex. Sometimes this complexity is the consequence of interacting thresholds; but that need not always be the case. But whether or not we have complete knowledge of the threshold, whether a point or a range is critical, whether the threshold is complex, a given threshold can none the less be passed when the actual practice is sufficiently dense.

Furthermore, our lack of knowledge about given thresholds should not be confused with another problem, namely, the fact that we ordinarily have very imprecise and incomplete knowledge of others' behaviour. We might have general information about a given threshold without knowing whether it will be passed and the effects produced. We might know the critical frequency for lawn-crossing in a certain case without also knowing the behavioural dispositions of others, how their acts can be influenced by example, and so on—in sum, without knowing whether, under the surrounding circumstances, the grass will suffer damage at all. We might know how many votes are required in a given case without knowing voting trends, how significant any particular vote will happen to be, whether the issue will pass after all. We shall find it important to make this distinction: between the facts of the case, the actual circumstances, and our knowledge of or beliefs about them.

Let us now more carefully attempt to account for the occurrence (that is, the production) of threshold effects from a causal point of view. We know that when the lawn is crossed infrequently the damage is negligible, but that frequent crossings cause appreciable damage. The several acts seem exactly alike; the lawn may be crossed in very much the same way in either case, within or outside the general practice, when the damage is or is not done. We know that a number of lies can be told, each of which entails some loss of confidence in the reliability of

communication, all of which may be in all obvious respects causally identical, and which have a total effect equal to the effects of the several lies taken separately. But if enough such lies are together told in the same way, then the total loss of confidence and other effects will not equal the sum of the effects produced by these lies taken separately. We know that one favourable vote will (ordinarily) be insignificant, but that all such favourable votes (or a critical number of them) will make the difference between election and defeat. But how is it that n acts that, when performed separately, each have effects **E**, when taken together produce total effects equal to something other than $n \times \mathbf{E}$? How can there be causal non-linearity?

Clearly, the 'taken together' is crucial. In providing a causal account of the occurrence of threshold effects, we have to make essential reference to the fact that the otherwise similar acts are performed *at* a certain density, for if they are not so performed, the threshold effects cannot (causally) be produced. Let us use the expression 'general practice (of acts of kind *B*)' to signify a practice (of acts of kind *B*) which is sufficiently dense to pass the threshold and hence to produce the associated threshold effects. We may thus say that a general practice (of acts of kind *B*) is a causally necessary condition of the occurrence of the threshold effects (associated with acts of kind *B*). When we say, perhaps misleadingly, that the same (sort of) acts when *taken together* produce effects that are different from the effects produced by an equal number of such acts *taken separately*, this complex causal relation which is now built into our notion of a 'general practice' is also implicit in the differences entailed by 'taken together' and 'taken separately'.

This is not a matter of how we view the acts. (Harrod perhaps made the mistake of thinking it was, in some sense.) This is not a matter of conceptually taking them together or separately. It is a difference of *circumstances* which may surround the performance of acts of a given type. The acts are performed at some determinate frequency or density. This is the special causal contiguity among *otherwise* similar acts which affects their causal efficacy. The pattern of others' behaviour is therefore a causally relevant circumstance of acts that have associated thresholds.

B. *Causal Linearity and the Relevance of Others' Behaviour*

The notion of a causal property (or a causal description) is conceptually linked with that of causal efficacy, the effects ascribable to acts. In the first place, a property cannot be said to be causally relevant unless something results (or would result) from the performance of an act in virtue of its having that determinate property—unless some difference in subsequent states of affairs is causally connected with the act's having that property instead of some other. Secondly, differences in the causal efficacy (in the kind or magnitude of effects) must be ascribed to differences in the acts themselves—that is, to differences that can be expressed in differences in descriptions of the acts (including the agents and circumstances) which go beyond merely describing the respective different effects *as* effects.

Accordingly, if seemingly similar acts are unequally efficacious, then we must infer the existence of some undisclosed causally relevant differences in their descriptions. In other words, acts are exactly similar in their causal descriptions if, and only if, they are equally efficacious. Such acts fall into *causally homogeneous* classes which are determined by their complete causal descriptions. This is the ideal towards which our classification of acts as causally similar is directed.

The notion of causal *homogeneity* thus incorporates the concepts of exact causal *similarity* and *equal efficacy*. This entails a connexion between causal homogeneity and causal *linearity*. Consider the latter concept anew. It is a characteristic of the tendency/effect relation for acts of a given description such that $\mathsf{T} = n \times \mathsf{E}$, where T is the total effect of the performance of every member of a class of n acts which could be (but will not necessarily be) performed, and where E is the effect of any member of the class. Non-linearity obtains, therefore, if T does not equal $n \times \mathsf{E}$. But we at once run into difficulty: what does E represent in a condition of non-linearity? It cannot unequivocally represent the effect of every and any member of the class whose tendency is T; for if that were the case, we should be faced with the impossible task of accounting for the difference between T and $n \times \mathsf{E}$. If the acts were equally efficacious, then in a condition of non-linearity some effects would be mysteriously appearing or disappearing.

And of course effects do not disappear (in this sense), nor can they just appear; they are not, for example, ascribable here to anything other than these *n* acts—or they do not belong in T. Thus, equal efficacy among the acts entails, and is entailed by, linearity; and unequal efficacy entails, and is entailed by, non-linearity.

Hence a condition of non-linearity can hold only for classes of acts not all of which are exactly causally similar in their effects and in their determinate properties (and descriptions). Non-linearity is a possible, and of course a very common, condition; this is precisely the point of Harrod's argument. But it is a condition of classes of acts which have been *incompletely* determined from a causal point of view. We may have good reason for grouping acts in such a way that non-linearity obtains; but none the less there will remain differences in their causal descriptions.

This argument perhaps suffers from a reliance upon rather vague notions of causal efficacy and relevance. But its point is quite simple. The intention is to put threshold effects on a level with all other effects of acts. They are effects of acts, after all; they happen to be the effects of acts of a certain character, of acts performed in certain kinds of circumstances, that is, as part of a general practice. However interesting and important this phenomenon may be, the special effects must somehow be ascribable to acts *within* the general practice. Thus, in addition to having their ordinary, run-of-the-mill effects, as it were, some acts also have *threshold-related* effects, effects that they have solely because they contribute to the production of threshold effects. A causally homogeneous class is one the members of which are exactly similar in respect of their threshold-related effects as well as any others. Non-linearity obtains, in particular, for classes of acts which are *mixed* in respect of whether or not the several acts have threshold-related effects and to what extent such effects may be ascribed to them. But when non-linearity obtains, that is a sure sign that there are relevant distinctions yet to be made among the descriptions of acts thus grouped together.

Our general line of argument also presupposes some doubtful propositions about the nature or proper characterization of acts. We are supposing, for example, that particular acts are discrete and identifiable and that, in some strong sense, the same concrete act may be described in various ways, more or less completely,

while there is, in principle (or ideally), some complete description of an act. Some difficulties which might appear here merely result from our special use of the terms 'act' and 'action', which comprehend features of the agent and circumstances as well as what is ordinarily considered to be the act proper. Other difficulties undoubtedly result from the fact that, among the alternative descriptions of supposedly one and the same act, some include at least part of what, on the basis of other descriptions, would be considered consequences of the act. But if there are irremediable difficulties here, I do not think they affect the present argument. For the present argument proceeds within a given tradition which has such questionable presuppositions regarding the way acts are to be viewed. In the present context, I am not criticizing utilitarianism (or ethical teleology) as such; nor am I primarily weighing the relative merits of two kinds of utilitarianism. I am rather comparing the substantive import of two kinds of principle, and in this respect I am working within this tradition and must accordingly accept for the present the requisite presuppositions. These may freely be questioned in another context, in a critique of the utilitarian programme as a whole.

But is it none the less unreasonable to suggest grouping acts in causally homogeneous classes, since this requires a complete causal description of acts—an impractical proposal? The argument does not suppose that we can actually carry out such a classificatory programme. The point is merely that non-linearity signals relevant differences. I shall suggest in greater detail just what kind of distinctions can be made among such acts. But would not such a classificatory scheme be objectionable in another respect, since the groupings would inevitably be reduced to unit classes (classes with only one member)? No, unit classes would not necessarily result. Since we are generally excluding spatial and temporal predicates, acts will not be relevantly different simply because they are not identical (because they are not one and the same particular act). But even if unit classes would result, this does not seem a sufficient reason against viewing acts in the way required by the argument. For if acts are relevantly different to that degree, we must admit it. We cannot reasonably ignore some relevant differences just for the sake of preferred class sizes. But of course this does not assuage the

feeling there may be something wrong with this result, that we should lose the point of the generalization test if it often happens that 'everyone's doing the same' is a vacuous condition.

I shall complete the causal argument after another parenthetical word. In the following I shall continue to consider lies, lawn-crossings, and voting as if they determined causally homogeneous classes—at least up to a critical point soon to be specified. (They will represent what I shall call causal similarity *in vacuo*.) I will use this shorthand even though the several acts that would be performed as part of such generically defined practices actually would be done in relevantly different ways, under relevantly different circumstances, and by relevantly different agents. It must also be remembered that in an adequate causal analysis of thresholds we would group acts together in respect of one threshold at a time. Such generically defined acts as lawn-crossings may have several associated thresholds, not all of the same type (though perhaps some are not related to the description of the act as given). But we shall simplify and speak of lawn-crossing only in respect of damage to the grass, and so on.

The arguments for non-equivalence must be taken in respect of acts that are completely relevantly described. Harrod suggests that he does so by indicating that the lies performed both within and outside a general practice are 'performed in precisely similar relevant circumstances' (p. 148). Harrison moves in this direction by seeking a complete and general criterion of relevance. But regardless of how we view the non-equivalence arguments or assumptions of Harrod and Harrison, it is clear that we must consider complete descriptions only. For if we view acts in respect of partial descriptions, we may obtain an erroneous impression of what a given principle implies about their rightness or wrongness. Thus we might mistakenly conclude that non-equivalence (or equivalence) obtained if we did not consider complete descriptions only—if, that is, we failed to consider some relevant features of some acts. Accordingly, only non-linearity of completely relevantly similar acts is of interest to us. But we have shown that causal non-linearity for exactly causally similar acts is impossible. Thus, the arguments for causal non-linearity must rest upon a failure to consider all the relevant causal factors. This failure is systematic; one kind of circumstance

has systematically been overlooked or excluded, namely, the pattern of others' behaviour. This is the crucial factor in the production of threshold effects, and our acknowledgement of it restores linearity. Since this particular issue is central to the argument, and since the normative argument rests upon the causal one, it will help to pursue the matter further.

Let me introduce another expression, an admittedly 'loaded' one: 'description *in vacuo*'. A description *in vacuo* '*D*' is a relevant description of an act which is complete up to a point: no part of it concerns the pattern of others' behaviour in respect of *D* (nor of part of '*D*' in respect of that part). In particular, no part of '*D*' concerns the relation of the given act to a general practice. In other words, '*D*' is the maximum relevant description that is common to acts both within and outside the general practice of *D*. Here we are of course using 'general practice of *D*' to signify a practice sufficiently dense to bring about those threshold effects associated with *D*. We are concerned here with causal descriptions *in vacuo*. (We can disregard acts that have no associated thresholds—for them there is no need for such distinctions; for them there are no problems with respect to linearity.) This notion will now be used to explicate the paradoxical aspect of Harrod's argument and the systematic incompleteness of description.

Thus, if it is claimed that the million lies performed *without* limit of time and space are performed in 'precisely similar relevant circumstances' to the million lies performed *within* certain such limits, and if this is to suggest that the two sub-classes of lies—those performed outside and within the general practice, respectively—have identical relevant descriptions, then clearly the description being entertained is only the description *in vacuo*. For the maximum common description (roughly 'lying') is incomplete for both sub-classes of lies. The fact that some of the lies are and some are not performed as part of a general practice is a causally relevant circumstance distinguishing the acts of one sub-class from those in the other. It is a causally relevant difference because general practices are causally necessary for the production of threshold effects. More precisely, a necessary condition of an act's having threshold-related effects is that it be performed as part of a general practice.

If we view the lies *in vacuo*, out of (this part of) their social context, then it may appear that acts of the two sub-classes are equally efficacious and similarly describable. But this results from viewing the acts only in respect of their non-threshold effects and properties; this, in turn, results from ignoring or determinedly excluding the relation of the particular acts to general practices and to thresholds. Hence the illusion that there can be a condition of non-linearity for completely described acts.

I shall deal below with reasons philosophers have had (or may have had) for not allowing consideration of the social context, or for allowing it only in a qualified way. Here I might suggest why the social context of acts sometimes seems to have been ignored entirely. In the first place, the significance of the social context could hardly have been noted without an analysis of threshold effects. Such an analysis has been discouraged by a failure to separate the causal and evaluative factors, and by the absence of a theory of relevance for utilitarian generalization. Secondly, in Harrod's case the descriptions *in vacuo* may have appeared complete for a special reason. It might have seemed that the distinguishing features of otherwise similar acts within or outside general practices are spatial and temporal location: witness his specification of what amounts to a general practice as determined by certain limits of time and space. But at the same time he might have been aware that spatial and temporal location cannot be relevant, that the extra (threshold) effects cannot be attributed to acts in virtue of such properties. Thus, the description *in vacuo* may have seemed as complete as possible.

It is doubtful, however, that the spatial and temporal specifications of acts are sufficient for distinguishing the two sub-classes, for determining a general practice, since the causal situation is far more complex than such reference can suggest. This is clearly so with regard to effects that turn upon interpersonal relations, as in the case in question, lying. Moreover, we shall also see that further distinctions can be made beyond mere participation in a general practice. But in any event, the social causal context satisfies the necessary conditions of causal relevance; the pattern of others' behaviour is the kind of circumstance in virtue of which the effects of some lies are enhanced and thus the threshold effects are produced.

One might alternatively claim that others' behaviour cannot be taken into account, on practical grounds. Perhaps we cannot consider such factors because we generally do not or cannot know enough about thresholds and about others' behaviour in relation to them, that is, about those factors which determine which subclass a particular act falls into. Part of the difficulty is that the social context may be in flux: whether or not a given act finally happens to be performed as part of a general practice may depend upon others' contemporaneous or future behaviour.

But such difficulties are no different in principle from those one faces in reckoning any causal circumstance of an act, especially circumstances involving others' behaviour, but not only of that kind, and not only in respect of others doing *the same*. All one can say is that, however difficult it may be to obtain knowledge of the facts, the facts determine the act's full causal character, its complete description, its proper classification.

Harrod's paradox dissolves. How can *n* lies produce more than *n* times the effect produced by one such lie, even when the several are 'performed in precisely similar relevant circumstances'? Only if the lies, however similar, are none the less causally different; only if the apparent causal homogeneity of the generic class is a homogeneity *in vacuo*.

It is interesting to note a shift in Harrod's argument. First, when suggesting references to threshold effects and when using these as the basis of his non-linearity argument, he treats relevance in effect as relevance *in vacuo*. But later (pp. 151–2) he explicitly allows consideration of others' behaviour. In the first stage he argues for 'refining' utilitarianism by shifting from effects to tendencies, from simple to generalized utilities. Later he introduces a 'second refining principle' that allows us to consider whether the practices dealt with in the first stage are very general—in effect, whether they produce threshold effects—and hence to determine whether the utility producible by their generality actually is being achieved. There is a virtue in separating stages in this way, for we may regard the consideration of others' behaviour, that is, to what extent others are doing the same, as a special sort of consideration, added to all the rest only when the description *in vacuo* has been determined. But Harrod's second refinement is actually implicit in the first, since the social context

may be viewed as a causal (or general utilitarian) circumstance on a level with any other. What Harrod failed to see was, first, the sweeping implications of the shift from effects to tendencies; secondly, the special implications of his second refinement. For, once we admit consideration of others' behaviour, we eliminate the divergence between the two kinds of utilitarianism, and hence we eliminate the point of the new kind of principle. Once we have a theory of causal and general utilitarian relevance, the relevance of such factors becomes clear; and the premium we are obliged to place upon completeness of description requires that in principle we *must* take such circumstances into account.

Performance as part of a general practice (of *D*) is a necessary condition of such acts having threshold-related effects. Therefore, acts which are performed outside the general practice are relevantly different in just that respect. There is no question of non-linearity for this sub-class of acts, since they have only non-threshold effects and properties. Thus, by making this distinction within the generic classes of acts (roughly lies, lawn-crossings, voting) in which non-linearity appears, we separate out two sub-classes, in at least one of which linearity is restored.

But matters are not so simple in the other sub-class, composed of acts which are both *D* and performed as part of the general practice of *D*. This sub-class is not in general causally homogeneous. For the term 'general practice' here signifies merely that the practice is sufficiently dense to produce threshold effects. This moderately restrictive usage captures the important connotations of the philosophical uses of the term, especially the condition that the total, cumulative effects of some kinds of acts are modified when the practice is general. But now we have introduced a more powerful concept, namely, 'threshold'. And we find that being performed outside a general practice is a sufficient but not a necessary condition of an act's having no threshold-related effects; for being performed as part of a general practice is a necessary but not a sufficient condition of an act's having threshold-related effects.

That is to say, some acts which may be said to be performed as part of a general practice do not always contribute to the production of threshold effects, and, since other acts within the sub-class of acts performed within the general practice do

contribute, that sub-class will not be homogeneous. For example, suppose there is a general practice that we can roughly describe as the frequent crossing of a given lawn (perhaps within a restricted period) such that damage occurs to the grass. Now some acts performed within this context (i.e. acts of crossing the same lawn during the period when the critical frequency is passed, and so on) would count as being part of the general practice even though they cannot be counted as contributing to the damage in question. Suppose that most of the crossings occur at one strategic corner (the crossing of which saves many people many steps) and that there, but only there, is the grass appreciably damaged (so that patches of soil show, and so on). But suppose also that Jones contemporaneously (and perhaps frequently) walks across the middle of the lawn, not across that corner. The middle of the lawn is not damaged since very few choose to cross there. There would be reason to consider Jones's act as part of the general practice—unless we further restrict the concept of a general practice—but however we would be inclined to treat Jones's act like the others in the practice, we must admit that his act simply cannot have threshold-related effects under the circumstances. In this limited respect, therefore, Jones's act is causally less akin to other acts within the general practice (at the corner) than it is to otherwise similar acts performed outside it. For these latter acts also have no threshold-related effects.

The importance of this sort of case stems from the fact that questions of utility—including generalized utility—turn upon questions of causal efficacy. No harm would come from *everyone's* doing what Jones does. That is, if everyone who could were to do that sort of thing (fully described), the lawn would not thereby be damaged. For part of the causal character of Jones's act are circumstances sufficient to preclude that such an act will contribute to any appreciable damage. And thus, though the general utilitarian may be inclined to treat Jones's act like the others in the general practice, to consider it no more justifiable than those others, it is clear that Jones's act must be regarded differently. For the tendency of the kind of act Jones performs is not undesirable, while the tendency of those others is.

The result is, not only must we distinguish between acts within

and outside general practices, but we must entertain finer distinctions, distinctions based upon finer details of the causal circumstances surrounding the particular acts, distinctions concerning the precise relations of the acts to the threshold in question. Some such distinctions will now be roughly sketched, but these will not nearly exhaust the possibilities.

Consider first, sequential acts; our example will be voting in a public roll-call. The case will be clearest if we suppose that the number m of favourable votes (for a given issue) is less than the total number n of votes cast and more than the minimum k required for passage. Here the general practice may be regarded as the collection of favourable votes cast. In the ideal case we can disregard special conditions and voter interactions which would cause differences among the favourable votes through the kth.

Thus the kth favourable vote crosses the threshold in question. When k favourable votes are cast, passage is assured according to the rules of the balloting, regardless of how subsequent votes are cast. Therefore, the favourable post-threshold votes (numbers $k+1$ through m) are as inconsequential with regard to this threshold effect as the unfavourable votes. The amassing of a substantial majority above the minimum required by law may admittedly be important, perhaps even for the future of the issue in question—but not for this threshold effect, not for this passage itself. The redundant favourable votes, those after the kth, are therefore like Jones's crossing of the damaged lawn: they seem in so many respects just like the other favourable votes and would most reasonably be counted as part of the 'general practice', but circumstances make a crucial difference.

In this simplified case we have assumed a constancy of circumstances up to the threshold, with the result that the first k favourable votes can be regarded as equally efficacious, as composing a causally homogeneous class each member of which can be ascribed $1/k$th the threshold effect. But in any actual case of public roll-call voting additional effects complicate matters. There are, for example, especially influential voters and the bandwagon effect (as well as its opposite, support for 'the underdog'). Such complications add nothing essentially new; some undoubtedly involve threshold effects

themselves; further factors of the same kind must merely be reckoned with.

In these complex cases the size of the causally homogeneous classes will decrease, for acts similar in one respect will differ in another. We must distinguish carefully between various kinds of effects—for example, between effects related to passage in the formal sense, in accumulating the legally requisite number of favourable votes, and effects related to passage in respect of influence, leadership, and so on. Each of the favourable votes might be performed in causally different circumstances (apart from the bare fact that, as each vote is cast, the circumstances involving the pattern of voting thus far is altered). This is especially so if we consider the motives of the voters and how they may weigh the probabilities of passage *vis-à-vis* their conflicting interests as circumstances of their actions. If one early voter is particularly influential, it might be reasonable to ascribe to him most of the credit for passage—though of course not all, for his leadership presupposes followers whose acts are necessary. There is also the case of the critical final vote that breaks a tie (as when the Vice President of the United States votes in the Senate). In some respects this vote is vastly more significant than the others; but again, not all the threshold effect can be attributed to this one act, since no redundant votes have been cast and each favourable vote cast is, under the circumstances, essential for passage.

The point is, that we may easily oversimplify and select too generic a description—ignoring certain relevant causal factors—when applying the generalization test. For while it is true that all of the *m* favourable votes are alike in that respect, as favourable, they also may be very different in other respects, cast from different motives that may prove to be causally relevant, at different junctures, and so on. When all the relevant factors are considered, it is likely that a vast spectrum of similarities and differences will be exposed.

For contrast, let us consider non-sequential acts; our example is a standing vote (a roughly simultaneous secret ballot would do as well). The same conditions hold as before: the number of favourable votes cast is less than the total but more than the number required for passage. We may also have to suppose for

simplicity that each voter decides independently, is unaware of the intentions of others, is not under party discipline, and so on. In this case, unlike the other, all the *m* favourable votes may most reasonably be viewed as composing a causally homogeneous class. For it does not seem possible to distinguish, among the *m* favourable votes simultaneously cast in these ideal conditions, some *k* that do (as opposed to some $m-k$ that do not) actually contribute to the production of the passage-threshold effect. We would have no way of deciding which particular acts have threshold-related effects and which do not, since the circumstances we have supposed guarantee, in effect, a uniform condition of voting. Thus all the acts within the general practice (the *m* favourable votes) are here alike and must be treated alike, despite the fact that fewer favourable votes would have been sufficient for passage.

This is the sort of case upon which Harrison's argument for non-equivalence rests; it is the sort of case which, if inadequately analysed, may lend spurious credit to the notion that it makes an essential difference, in threshold-related cases, to consider acts as members of classes of similar acts instead of one at a time. Let us therefore examine this case and Harrison's position more carefully.

Harrison avoids the paradoxical results of Harrod's argument by acknowledging that acts within the classes under consideration (the generic classes, described *in vacuo*, e.g. lies) are *un*equally efficacious. For he wants to account for the extra (threshold) effects and is also aware that they must be attributable to acts and to nothing else. He therefore argues:

> Actions of the class in question [i.e. any acts having associated thresholds] must be so related to one another that, if they are not performed in the majority of cases, then they will not produce good consequences [or bad consequences, *mutatis mutandis*]—or, at any rate, not such good consequences—in any. They must be related to one another in such a way that the good consequences produced by those of them which do produce good consequences are dependent upon a sufficient number of those of them which do not have good consequences being performed. (p. 120.)

Here Harrison comes closest to providing an analysis of threshold effects. But his analysis is not sufficiently general and, most important, it is not integrated with a theory of relevance. It suffers from one further defect. He wishes to claim that there are differences in the causal, and indeed the utilitarian, efficacy of acts within the generic classes, but to deny that further relevant distinctions can be made within them. Or at least he is not aware that, once making these further distinctions, we are obliged to treat the various sub-classes according to their own respective generalized utilities, and not according to the generalized utility of the generic class. For example, those acts, within a general practice that has undesirable threshold effects and therefore causes harm on the whole, that do not cause harm because they themselves do not have threshold-related effects, are acts which can justifiably be performed on either simple or general utilitarian grounds. But the proponents of utilitarian generalization and of rule-utilitarianism want to condemn all these acts within the general practice regardless of whether they themselves are harmful. Such a blanket condemnation is assumed to be the point of the generalization test and it is the philosophic objective at hand.

Here is Harrison's error in detail. He argues (pp. 120–1) that some of the acts within the general practice do and some do not contribute to the production of the threshold-dependent value and presumably, therefore, to the production of the threshold-dependent effects themselves. This makes plausible his claim that there is a qualitative difference between the simple and the generalized utilities of such acts. (This difference would hold, of course, for those acts without threshold-related effects, but Harrison suggests that it holds for all.) Now Harrison argues that the several acts within the general practice must be treated the same, even though they have different effects and utilities, *because* the performance of some which do not actually contribute to the production of threshold effects (cf. 'a sufficient number of those of them which do not have good consequences') is *necessary* for the production of those effects (i.e. 'the good consequences produced by those of them which do produce good consequences are dependent upon' the performance of the others). But wherein lies their necessity? Whichever acts are in fact causally necessary for the production of the threshold effects (given their determi-

nate circumstances) should be accorded credit, as it were, for their part in its production; some threshold-related effects must be accorded all such acts. But if no such effects can be ascribed to them, then it is untenable to claim that those actions are necessary. Thus the sub-classes must not only be distinguished, but also treated differently, for acts that are not causally related to the production of the threshold effects are to that extent indifferent, while those contributing to the production of value-laden threshold effects are by no means indifferent.

However, Harrison's argument may gain credence from the examples of non-sequential acts, such as those in our second voting case, which is similar to one of his examples. In that case we can see the plausibility of the notion that some acts which are part of the general practice (some of the *m* favourable votes) are indifferent and yet should be treated the same as the others. *m*—*k* votes are unnecessary for passage; but since all the *m* favourable votes are cast in exactly similar circumstances (we are supposing), no particular sub-class of favourable votes can be selected as actually producing or not producing the passage-threshold effect, and hence all *m* votes must be treated the same and similarly described.

But it is essential not to confuse two things: (1) the fact that *k* favourable votes would have been sufficient, and therefore that, if only *k* favourable votes had been cast, the threshold effects would be ascribable only to them, with (2) the claim that, under these circumstances, some of the *m* favourable votes do not actually contribute to the threshold effects. If it takes six men to push a car up a hill and, not knowing this, eight lend a hand and do the job, what are we to say? If all actually pushed, and pushed equally hard, and delivered equal forces, are we to say that only some of them actually contributed to the effects because fewer *could* have done the job? Indeed, does it make sense to propose a division into sub-classes unless there is in principle a criterion for making the distinction in such a case? On the other hand, if circumstances are such that one could, in principle, relevantly distinguish sub-classes within the general practice, then we should be obliged to infer that the votes that do not actually contribute to passage need not have been cast, that these votes should be treated differently. This is the conclusion that the generalizers apparently want to avoid.

A variation on the argument we have attributed to Harrison is also worth noting and answering. It is sometimes argued, for example, that following Act-Utilitarianism could have disastrous results, in the following manner. Since I know that k votes are required for my candidate's election, and since I am also quite certain that altogether m votes will be cast for him (or at any rate enough for election without my own), I can infer that my vote is insignificant, for it will make no difference to the outcome if I fail to vote for him. (I could even vote against him, if the margin is wide enough.) Now some want to argue that despite these facts one none the less has a duty to vote and that this duty has utilitarian grounds. For what would happen if everyone did the same? The answer given is, 'if all his other supporters were to do as I do, and fail to go to the polls, the man would not be elected' (Harrison, p. 107).

But it would be impossible for *all* his other supporters to do what I propose to do—if one takes all the relevant circumstances into account. If everyone (who can) does what I propose to do, the man will be elected anyway. For I propose, not simply to refrain from voting for him, but rather to refrain from voting for him when in fact my vote is inconsequential. That my vote is inconsequential is established by the conditions supposed at the beginning. These are the conditions which show that there is a qualitative difference between the simple utility of my act and the generalized utility of the generic class of acts; but they are also the conditions that require our distinguishing subclasses within the generic class. And we cannot use these conditions at whim, granting them now, denying them (in effect) when it comes to generalizing and classifying the act in question.

This retort to one persistent form of the non-equivalence argument may be obscured by many factors. For one thing, the character of the act is usually inadequately specified to begin with. Perhaps the point at issue is voting (in general); the point the generalizer wishes to make concerns voting (without qualification), without regard for special circumstances (except perhaps very great hardship cases). So one does not bother to integrate the fact that at least k favourable votes will be cast regardless; and that it is impossible for this to be a circumstance surrounding more than $m-k$ occasions for voting; and that, if $m-k$ supporters

abstained, their candidate would still be elected; and that all these things are of causal relevance—and, as we shall presently see, also of general utilitarian relevance.

Secondly, this retort can be obscured by the formulation of the original argument in the probabilistic mood (which is suggested in our own formulation). Thus Harrison says, 'I do not think that the addition of my vote to the total number of votes which are cast for him is going to make any difference to the result of the election' (p. 107). And it is conceivable that all his supporters should think the same—should even have good reason for believing the same—and thus fail to vote. That is, if everyone, who merely had good reason to believe that enough others would vote, failed to vote, the unwanted and avoidable results could be produced. But of course if we make the condition stronger, if we say 'I *know* that enough others will vote for him', then this rejoinder does not hold. It should be observed, now, that I have not concerned myself with beliefs about or probabilities of effects occurring. I have considered only the actual facts of cases, the actual circumstances surrounding the several acts. The question is, given the *real* circumstances surrounding this act, should it be performed? (Or in other words, what is the actual implication of this non-probabilistic principle?)

There is a residue of difficulty here, occasioned by widespread disagreement and indecision expressed by many writers on the subject. I shall expand below on the relevance of facts, knowledge, and beliefs of others' behaviour, from a utilitarian point of view.

c. *The General Utilitarian Relevance of Others' Behaviour*

Causal linearity obtains within classes of acts the members of which are exactly similar in all causal respects. This conclusion might be translated directly into the normative sphere to establish utilitarian linearity. General utilitarian properties are a subclass of causal properties, and utilitarian relevance is, presumably, similarly tied to causal relevance. But there is a disanalogy between causal and general utilitarian properties. While these are of course both properties of particular acts, the former are related to the effects of particular acts, whereas the latter are grounded upon generalized utilities, upon the values of the

tendencies of acts, the effects of everyone's doing the same. Since there is this disanalogy, an argument is required for utilitarian linearity.

Narrowly speaking, our question is whether non-linearity (meaning, hereafter, utilitarian non-linearity) is a consequence of threshold effects. For in no other context do we have reason to suppose non-linearity; in every other context the results of everyone's doing the same are simply related and obviously reducible to the simple utilities involved. In every other context, in the absence of threshold relations, it seems indifferent whether we calculate utilities the simple or the generalized way. Accordingly, in the following argument for linearity I consider threshold relations only. But I shall also present an argument having no such restrictions.

First I shall sketch a constructive argument for linearity. Suppose that non-linearity appears to obtain for acts *D*. The description '*D*' is of general utilitarian relevance so far as it goes, but it is not necessarily complete. Our question is, whether it is necessarily *in*complete. We assume, within the present context, that '*D*' is a relevant description *in vacuo*: in the basic cases, '*D*' includes no mention of general practices or of threshold relations. Thus, we exclude the case where one part of '*D*'—say '*B*'—consists of specifications concerning to what extent others are doing *A* (where '*A*' is another part of '*D*'). '*D*' is thus the sort of description regarding which questions of linearity seem to arise; it is a description that may seem complete, but it has an associated threshold and lacks mention of threshold-related circumstances.

By adding further relevant specifications to '*D*' we can restore conditions of linearity. I propose to add specifications that we know to be causally connected with just those utilities giving rise to doubts about linearity, that is, with the value-laden threshold effects of the general practice of *D*. These specifications serve to mark off acts that can, or do, have threshold-related utilities from otherwise similar acts that cannot or do not. And they mark off such acts, separating them into different subclasses of *D*, by means of referring to circumstances related to whether the several acts produce the effects that have the value. Thus these candidates satisfy the requisites for general utilitarian descriptions.

Consider the most general case, 'performed within the general practice of D', i.e. performed within a practice sufficiently dense to produce the threshold-dependent value associated with D. Let 'C' stand for this predicate and '$\bar{C}$' for its negation. Each of the n acts will be either C or $\bar{C}$; where 'C' is not true of an act, of course '$\bar{C}$' is true of it. In the most general case, 'C' will not be true of every one of the n acts; but even if the threshold-dependent value is not produced, 'C' will be true of a sufficient number of acts which could be performed but are not performed. Note that all the acts under consideration are acts which could be performed (occasions for doing D); they are not necessarily performed; but on the other hand some of them are (probably) performed. 'C' will not be true of any performed acts unless there is actually a general practice and the value is produced; if there is no general practice, then all performed instances of D will be $\bar{C}$. In any event, '$\bar{C}$' will be true of acts D that are outside the general practice; such as when one votes for a losing side, or breaks promises when these are generally kept, or helps try to push a car which refuses to move anyway, and so on. These descriptions, when respectively true of acts, will always be relevant. For $\bar{C}$ is a sufficient condition of an act's having no threshold-related utility; threshold-dependent value is attributable only to acts that are C as well as D. Thus if everyone who could did DC (a stronger condition than doing D), the threshold-dependent value would be produced, since every act within the general practice would be performed, and thus there would necessarily be a practice sufficient to produce that value. Whereas if everyone who could did $D\bar{C}$, the value would not be produced—that is to say, it would not be produced by *these* acts, in virtue of *their* performance; if it were produced it would only be produced by DC acts, and it would not be produced at all if only $D\bar{C}$ acts were performed. The production of such value or lack thereof depends upon just these differentiae, C and $\bar{C}$; thus 'C' and '$\bar{C}$' are further relevant specifications of instances of D.

Linearity obtains for the sub-class $D\bar{C}$ because threshold relations are not involved there. That is, unlike acts D, acts $D\bar{C}$ are such that, no matter how frequently they are performed (within the limits imposed by the condition $\bar{C}$), threshold-dependent value cannot be produced by their performance.

Thus, the conditions leading us to suppose non-linearity for D are absent for $D\bar{C}$.

But being performed within a general practice, C, is not a sufficient condition of an act's having threshold-related effects. It follows that it is not a sufficient condition of an act's having threshold-related utility. Hence linearity does not necessarily obtain for the class DC; this class can often relevantly be further broken down.

Those DC acts that are not strictly necessary to the production of the threshold effects, and thereby to the production of the threshold-dependent value, and which therefore have no threshold-related utility, can be distinguished from the other DC acts by reference to the kinds of special circumstances already suggested in our discussion of causal linearity. We can simply proceed by further differentiating acts by reference to aspects of their circumstances (related to how others are acting) which are causally connected with the degree of value-laden threshold-related effects the acts produce. We should in this manner arrive (ideally) at a set of *teleologically homogeneous* sub-classes of D, the members of which are exactly similar to other members of the same sub-class in respect of their general utilitarian descriptions, their generalized and their simple utilities. And in all such classes linearity would of course prevail.

Suppose that we take (roughly) lawn-crossing as D. The foregoing argument tends to show that, if lawn-crossing is a utilitarian description *in vacuo*, then there will be linearity for such acts performed outside a general practice of lawn-crossing, i.e. for $D\bar{C}$. Moreover, although those acts within the general practice may not be exactly similar, none the less the class DC can be broken down so that lawn-crossings which actually contribute to appreciable grass damage (threshold disutility) are separated from those which do not.

Suppose, however, that there is no threshold for damage to the lawn. Suppose that, so far as the state of the grass is concerned, acts performed within and outside a general practice that is defined in terms of threshold-dependent value (and not in terms of effects as such) are equally efficacious. No matter how many times the lawn is crossed, the total damage of any sequence of lawn-crossings is (ideally) always proportional to the damage

caused by any one act that can arbitrarily be selected from any such sequence. In other words, there is always only an orderly, step-wise linear relation between the total effects of a sequence of lawn-crossings and the number of performances. This sort of case does not seem to fit into our general pattern of analysis. For we have treated the threshold phenomenon as basically causal. Here, however, it appears that the threshold is evaluative: the threshold is located in the ascription of value to (or possession of value by) the various states of affairs which are produced. For example, the disutility of a large degree of damage is simply out of proportion to the disutility of a smaller degree of damage.

In order to complete the argument for linearity we must take account of such cases and assimilate them to the general pattern of analysis. There are, I think, two ways in which a threshold can be evaluative in the respect outlined above. The first possibility develops as follows. There is no threshold in grass damage (or lawn damage), but some point in the accumulation of that damage is reached (when lawn-crossing is frequent to a certain degree) where the grass is damaged 'to such an extent as to detract from' somebody's 'pleasure in contemplating it'. In this sort of case, we can understand the thresholds in causal terms even though they are not a feature of grass damage. The thresholds in question are psycho-physiological; they concern human discrimination, our dispositions to pleasure (or to pleasurable reactions), and so on. This is assimilable to our general analysis, within our broad conception of causation.

There is another possibility. Suppose now that various states or conditions of the grass themselves have intrinsic value. That is, suppose that good grass is desirable, not just for the sake of our pleasure, as being instrumental towards it, but intrinsically. And suppose further that there are thresholds for the intrinsic value of states or conditions of the grass. For example, if we use some non-evaluative scale of grass damage as our standard, we might find that the desirability (intrinsic value) of the grass with less damage is disproportionately greater than the desirability of the grass with greater damage. We might understand this in the following way. Take two sequences of lawn-crossings with identical initial conditions, the several acts of which are equally efficacious (in terms of grass damage). One sequence is longer

(more dense) than the other. Therefore a different total damage occurs in each case, but the total damage is proportional to the damage caused by an act arbitrarily selected from either sequence. We find, however, that the grass which is less damaged (which has suffered the shorter sequence of crossings) has an intrinsic value disproportionately greater than the intrinsic value of the grass which is damaged more, taking as the basis for comparison the respective numbers of equally damaging acts.

In such a case, it might appear that equally efficacious acts have unequal disutilities. But if they are equally efficacious, they must have similar causal descriptions. Hence their general utilitarian descriptions cannot differ—for general utilitarian differences are simply a species of causal differences. But if these acts have similar general utilitarian descriptions and also different simple utilities, there must be non-linearity. And yet it seems that further relevant differentiation among the acts cannot be made.

But the very conditions of our example (and I suggest of any such example) lay the basis for further relevant distinctions. For the supposedly equally causally efficacious acts with similar causal descriptions but different simple utilities differ in this respect: they are respectively performed as parts of different practices (or sequences of acts) the members of which, while otherwise similar, are different just in that respect. They are performed in relevantly different contexts. We could, if necessary, as we sometimes do, define the differences between the practices in terms of differences in the results: in one case the total effects are of one degree; in the other case the total effects are of another degree. Those acts which contribute to the total effects of the greater degree *thereby* produce the greater utility; for the greater degree of damage is disproportionately less valuable just because of its greater degree. Such distinctions could be shown to be of general utilitarian relevance. But I shall not argue the point further, because the solution to a second problem contains an argument for linearity that is completely general, and covers this case.

The second problem is, how do we know that our original constructive argument for adding further relevant descriptions to acts is completable in the sense that a further general utilitarian

distinction is always available when required to restore linearity? Is it not possible that some pockets of non-linearity will remain, owing, not to limitations in our knowledge, but to the facts of the case? I shall argue that this is not possible.

Suppose we have a class of acts A which seem exactly similar in all general utilitarian respects and thus have one commonly assignable generalized utility, **G**. But it appears that non-linearity obtains within this class. What does this entail? Linearity obtains if, and only if, $\mathbf{G} = n \times \mathbf{S}$, where **G** is the generalized utility associated with a given complete general utilitarian description, i.e. it is the value that would be produced if every one of the n acts of that kind which could be performed were performed; and where **S** is the simple utility of any—hence of each and every—such act. Suppose there is one such act for which this **G**/**S** relation fails. (In the extreme case (Harrison's), the simple and generalized utilities that seem assignable to the act would differ in quality.) Now if this **G**/**S** relation fails for one act, the several acts A cannot have identical simple utilities. For if there were non-linearity, if **G** did not equal $n \times \mathbf{S}$, and yet the simple utilities of the several acts were the same, then there would be some unaccountable gain or loss of utility consequent upon everyone's doing A, a gain or loss equal to the difference between **G** and $n \times \mathbf{S}$. But of course this is impossible, for **G** is completely ascribable back to the several acts that would produce it if they were all performed, and the total value produced by the several acts must add up to **G**. This holds in every case, whether or not the descriptions are complete.

Let us consider one of the simplest possibilities (complex cases differing only in detail). Suppose that there are two sub-classes within A which may be identified as A_1 and A_2, distinguished now, not by reference to general utilitarian differences, but rather by the fact that every A_1 act has a simple utility equal to $\mathbf{S}_1$ whereas every A_2 act has a different simple utility equal to $\mathbf{S}_2$. Now if acts differ in their utilitarian efficacy there must be some differences in their effects (viewed non-evaluatively) to account for this, and the differences must be attributable to some difference in their properties. There must be some determinate property B, say, which is not mentioned in the description 'A' (lest 'A' be false of some of these acts), that is common to A_1 acts

but is not a property of A_2 acts such that, in virtue of their having this property (and the A_2 acts *not* having it), the A_1 acts each have simple utility S_1 (instead of S_2). The properties B and $\overline{B}$ therefore have some kind of utilitarian significance. The point at issue here is, whether such a property is necessarily a general utilitarian one.

The two sub-classes may now be referred to as AB and $A\overline{B}$. I would hold that distinguishing the two sub-classes in this way is of general utilitarian relevance. For the difference between S_1 and S_2 is causally related to the difference between B and $\overline{B}$. The condition of relevance is satisfied in the particular case; hence it is satisfied in the general case as well. A given description is relevant for the application of utilitarian generalization only if the universal performance of acts of the kind mentioned would produce some utility or disutility in virtue of the several acts having the property mentioned in the description. Now in any particular performance of an AB act, the act has some utility or disutility in virtue of the fact that the act is B (as well as A) instead of $\overline{B}$. In any general—or universal—performance of AB acts, as opposed to $A\overline{B}$ acts, part of the total utility produced is produced in virtue of the fact that the several acts are B (as well as A) instead of $\overline{B}$. Hence B and $\overline{B}$ are general utilitarian properties.

In other words, any differences in the simple utilities of acts can be ascribed, in effect, to differences in their general utilitarian character—and vice versa; this follows from a generalization of the foregoing sample argument. Hence non-linearity—which requires differences among the simple utilities of acts which are grouped together as similar—entails incompleteness of general utilitarian description.

One might object that this approach makes the notion of general utilitarian relevance too loose, its application too inclusive. Perhaps simple and general utilitarian properties should be mutually exclusive, the former related only to the simple utilities of acts, the latter understood, for example, as only those properties of acts in virtue of which the universal performance, and only the universal performance, of acts of those kinds produces some utility or disutility. This would restrict the sphere of general utilitarian relevance to threshold relations.

One difficulty we would face in thus restricting general utilitarian relevance is that this approach threatens to make the notion too narrow. If we press hard upon the requirement of universal performance, we eliminate the general case, that in which the special utility is producible by a general but none the less *non*-universal practice. This difficulty could be skirted, however, by explicitly reformulating the notion of general utilitarian relevance in terms of threshold relations.

But there are more serious difficulties. In the first place, we must allow room for consideration of non-threshold utilities (and therefore of non-threshold properties) in the application of utilitarian generalization, which this proposal would exclude. For non-threshold considerations can outweigh threshold-dependent ones. If we followed this proposal we would overlook or simply disregard many exceptions to rules that could be defended, on the more inclusive interpretation of general utilitarian relevance, on general utilitarian grounds; we would overlook what Harrison calls 'benevolent acts' as opposed to 'just acts', and related considerations. And note that nothing in the formulations or our analysis of utilitarian generalization suggests that we must restrict our attention to threshold relations and to threshold-dependent utilities. In applying utilitarian generalization we are simply obliged to connect the rightness and wrongness of acts with the good and the evil, respectively, which is producible by general practices. And we are presumably concerned with the total character of general practices—with their threshold and non-threshold aspects, and not with simply one aspect of them.

But the main theoretical obstacle to this proposed restriction of general utilitarian relevance is that all utilities are ultimately reducible to simple utilities; all utilities are ultimately ascribable back to the particular acts producing them. (This is obviously related to the fact that tendencies can be analysed into the effects of several acts.) For utilities are, after all, merely the value-laden effects of acts. It is immaterial that some utilities can be produced only by the performance of a number of similar acts; such utilities (as we have seen) are then ascribable to acts having this special character. Thus, generalized utilities are strictly reducible to simple utilities; and therefore general utilitarian properties are

understandable in terms of, and in effect are reducible to, simple utilitarian properties. One could not effectively separate simple and general utilitarian properties, except in name.

But is this to forsake the point of utilitarian generalization and of the generalization test by denying their special character?—Until we examine the connexion of utility to fairness, we can merely say that the generalization test can have at least two interpretations. The one we have given it so far is related to utilitarian generalization as a strictly utilitarian kind of principle. Now utilitarian generalization has been intended by some to be a means of accounting for considerations of justice or fairness on the basis of utility. (Witness Harrison's attempt to ground rules and duties 'of justice' upon (GU); see, e.g., pp. 108–12, 121–2, 125–8.) But whatever hopes or illusions may have surrounded these principles and the associated version of the generalization test, these must be sharply distinguished from the means adopted to carry out the programme. The programme has involved, in this context, merely a shift in viewing acts from their effects and their simple utilities to their tendencies and their generalized utilities. This is merely a change in the manner of reckoning utilities. No matter how one reckons, if one's sums are complete, the answers are always the same. Thus, the fact that general utilitarian relevance is so related to simple utilitarian relevance should not be surprising; for both kinds of relevance are, in the last analysis, fundamentally the same—just as both kinds of utilitarianism are fundamentally the same.

But, far-reaching as these general conclusions may be, they should not obscure the following point. The key to equivalence is a full inclusion of circumstances involving others' behaviour. For the omission of facts about others' doing otherwise similar acts from descriptions and generalizations is the source of the illusion of non-linearity; completeness of description restores linearity; and (as I shall argue) linearity precludes non-equivalence. In broad outline this traces the logical connexions between equivalence and others' behaviour.

One could conceivably argue for non-equivalence, therefore, by arguing that some crucial consideration of others' behaviour is irrelevant. This could be done—but no one in fact argues in

that way. As I have already suggested, it is not clear that the connexions between equivalence and others' behaviour have before been noted. It would seem, rather, that the question of equivalence itself has never closely been examined, and that, lacking a theory of relevance, lacking a sense of the premium to be placed upon completeness of description, preoccupied with certain given, socially or philosophically prominent modes of description, writers have found reason to suppose that the two forms of utilitarianism diverge.

Nevertheless, most writers on utilitarian generalization have some doubts about the propriety of taking others' behaviour into account, and some have offered arguments to the effect that such considerations cannot be unrestrictedly allowed. I shall here examine those reservations for which arguments have been given or that suggest some kind of argument, in an attempt to remove any obstacles remaining in the way of a full consideration of the social context of acts.

There are two kinds of moral argument against allowing full consideration of others' behaviour. One, the argument from fairness, does not deny that facts about others' behaviour are of general utilitarian relevance; but it denies the moral acceptability of the results of allowing full consideration of such circumstances when applying a form of utilitarian generalization. That is to say, certain implications of these principles that rest precisely upon a consideration of others' behaviour are inconsistent with the requirements of fairness. But since this argument presupposes that such facts are relevant, and thus that qualifications must be imposed externally, we may defer discussion of it until later (Chapter V).

Another kind of moral argument attacks the question of relevance directly. In discussing the nature of relevance, I have already alluded to this approach. It proceeds from the premiss that questions of relevance for a given substantive principle are settled by appeal to general, independent moral criteria or principles. Thus, when the question is asked, 'How are relevant similarities and dissimilarities determined?' the answer, on this line of reasoning, is, 'By consideration of what one is *morally* allowed or required to take into account.'

Thus if one asks whether *C* (e.g. that others are—or are not—

generally doing the same) is relevant, whether this circumstance can be considered, on this line of reasoning one should appeal to general criteria of 'moral relevance'. This approach is methodologically unsound and leads to a dead end. After all, how does one determine the criteria of moral relevance in general? To what substantive criterion might one appeal? It would seem that relevance should be and can only be determined by reference to the nature and content of the substantive principles in question. My earlier discussion of general utilitarian relevance was an attempt to work out this approach in the present case.

In his discussion of 'state of nature' conditions (pp. 152–61), Marcus Singer allows consideration of others' behaviour; the fact that others are or are not acting in certain ways is taken by Singer as sufficient for justifying one in falling back upon one's 'right of self-defence'. Singer argues:

> That certain other people are acting or may reasonably be expected to act in certain ways is part of the context in which the generalization argument is applied. To put it another way, that certain other people are acting in ways in which it would be undesirable for everyone to act, in ways that are generally [i.e. prima facie] wrong, is part, and an essential part, of the circumstances in which one is acting. (pp. 155–6.)

And yet it appears that Singer fails to appreciate the full significance of this admission. For he also says:

> It is simply irrelevant to reply [to the generalization argument], 'Not everyone *will* do it.' It is irrelevant because the argument does not imply or presuppose that everyone will. (p. 90.)

It would seem then, that Singer draws an apparently arbitrary line between others acting in ways which are wrong, other things being equal, and others not acting in such ways. But why should one be taken as part of the context of an act and the other rejected?

This aspect of Singer's discussion may be taken in the following way. The supposition that everyone will do the same should not be confused with a report or a description of the circumstances. The generalization test is hypothetical. The reply 'Not everyone *will* do it' here is taken as representing a misapprehension of the nature of the test or of the principles appealed to.

Of course it may also be a blunt rejection of the generalization form of moral argument. This interpretation of Singer's argument is borne out by its context, for here (pp. 90–95) Singer is introducing his discussion of C. D. Broad's rejection of such hypotheses (a rejection which we shall presently discuss).

But our earlier suspicions are intensified as Singer goes much further and makes the much stronger claim that

> the fact that not everyone will act in a certain way is irrelevant to the question whether it is right or wrong to act in that way. It is not a valid objection to the generalization argument, nor can it ever justify anyone in acting in the way in question. For the argument does not imply that everyone will act in that way, nor is this assumed in its application; and if this fact, that not everyone will act in that way, could serve as a justification, it would justify everyone in acting in any way whatsoever. (p. 145.)

But now consider this very common retort to the generalization test. Someone objects to my doing an act with a bad tendency by saying 'What if everyone lied (broke promises; failed to pay his taxes; &c.)?' And I reply 'But of course not everyone *will* do it.' This response might be taken (and meant) in a number of ways. I might simply reject outright an argument based on mere suppositions; I insist that you notice that your supposition is counterfactual as a way of getting you to give it up too. On the other hand, I might be elliptically suggesting, not merely that not everyone will do the same, but that too few will do it to make much of a bother. (In the extreme case, I should say merely that not everyone will do it when a universal practice is essential to the threshold effects.) That is, the practices of veracity (or communication), of promising, of military service, taxation, and so on, are not much affected over-all by a limited number of evasions, exceptions, violations, contraventions. I am claiming, in other words, that most people won't do what I have done or propose to do, i.e. that the general practice in question (the one hypothesized) does not and will not exist. Therefore, given these circumstances, an argument based upon threshold considerations has no force, for if not everyone who could do so (i.e. under such circumstances) were to do what I have done or propose to do

the undesirable threshold effects would not be produced. This is not a rejection of the generalization test, but an attempt to internalize relevant factors which have been overlooked.

One of the relevant circumstances surrounding my act is the fact that the evil-producing general practice does not occur. And this fact can be used to rebut the original presumption against my act, for my act was not completely relevantly described. Analogous things can be said of course about presumptions in favour of acts the general practice of which produces desirable threshold effects. When the act in question is performed outside the general practice, and thus cannot possibly contribute to the threshold effects, the generalization form of argument can have no special force. And this is not because we reject the process of generalization, but because the additional circumstances specified serve to mark off a class of acts every one of which is immune to threshold-related arguments. To admit this is not at all to 'justify everyone in acting in any way whatsoever'.

But there is another condition which may seem to suggest this. Consider the alternative response which I might make to your generalization test, especially when the act I propose (which is not fully specified) would have desirable effects, the more so when these are desired by me. I then say 'Well if everyone else were to do it, I'd be a fool not to do so as well.' There would be some reason in this seeming amorality. For if everyone else did the act in question and thus a general practice existed which produced the undesirable threshold effects, it would be irrational (from a utilitarian point of view) for me to refrain from performing the act. (If the threshold effects will be produced anyway, this case is similar to the sequential-act model in that my act will not necessarily be ascribed threshold-related disutility simply because it becomes part of a general practice.) And this holds when one adopts a general utilitarian point of view in particular; for no threshold-dependent harm would result if everyone who is in the agent's assumed position did the kind of act in question. Therefore, one can be *justified* in doing the act on general utilitarian grounds, precisely because the practice is general anyway, and despite the fact that it seems as if we are thereby allowing the universal performance of acts of which merely the general performance would yield bad effects. But

this is because we are taking the existence of the general practice as a standing condition of my act.

Let us now consider Broad's argument directly. He claimed that,

> so long as we believe that probable consequences are relevant in deciding the rightness or wrongness of an action, the particular circumstances under which the action is to be performed must be taken into account, since its probable consequence will largely be determined by them. And a very important circumstance must be the question whether other people are or are not going to do similar actions. (p. 378.)

In the usual employment of the generalization test, however, the supposition made is, as Broad observes, most often a counterfactual one.

> For, in practically every case where we consider what would happen if everybody acted as we propose to act, we know as surely as we can know anything that is not *a priori*, that by no means everybody will act in this way. (pp. 377–8.)

And Broad concludes that,

> if probable consequences are to be considered at all, we cannot and ought not to be guided by a false account of the circumstances; and the hypothesis whose consequence we are asked to consider in the method of false universalisation [i.e. the generalization test] is admittedly a false account of the circumstances in which our proposed action would take place. (p. 378.)

Thus Broad claims that a utilitarian argument must reckon in all the relevant circumstances, among which is others' behaviour. But he suggests only one kind of circumstance (that the practice is not universal, perhaps not even general) and also claims that the generalization test cannot acknowledge a fact of this kind—indeed he claims that the actual situation is misrepresented.

Broad's complaints are unfounded. If the test is applied to an act in respect of its complete relevant description, no 'false account of the circumstances' is employed from a utilitarian point of view. Or perhaps we should say that a false (counterfactual) supposition can be made in some cases, but not in any

relevant and therefore not in any worrisome way—not in any way which makes a difference. For by taking into account others' behaviour as it is relevant, we locate the act in question in its proper sub-class ('proper' from a utilitarian point of view), and it is the sub-class to which the act belongs that, with its associated tendency, is of significance for the rightness or wrongness of the act.

How, it may be asked, can the supposition be false with no consequent difficulty? Suppose that when I apply the generalization test and ask what would happen if everyone did the same, I hypothesize that every act that might be performed completely relevantly similar to mine *is* (or will be) performed. I might suppose, for example, that everyone who could do so (who will be in a position to do such a thing) will cross the lawn when others are not generally crossing it. For if there is not in fact a general practice of lawn-crossing, there may be occasion for several of us to cross the lawn without doing damage. But it may turn out that I am the only person among these who crosses the lawn, and that others who could have done so, without causing damage, refrain. Thus, my supposition that everyone in relevantly similar circumstances will do the same as I propose to do turns out to be false. But this makes absolutely no difference. For the point is, that the lawn is not generally crossed, and my act has been properly described and therefore properly assessed from a utilitarian point of view.

It is possible, of course, to act on a relevantly incomplete or incorrect account of the circumstances. This will involve making the misleading kind of counterfactual supposition with which Broad was perhaps primarily concerned. Thus, I might be wrong in supposing that others are not or will not be generally crossing the lawn. Upon making such an incorrect supposition, I might cross the lawn when it happens that others (or others with me) *do* generally cross the lawn, thus inflicting damage upon it. And such a false supposition can make a difference, if it leads me to think that my act will not have threshold-related disutility when in fact it does. But the difficulty arises not simply because the supposition made is counterfactual; it arises because my act has been *mis*described and therefore put into the wrong sub-class and assessed incorrectly.

Note moreover that, even if we lack a complete description but are aware of relevant facts about others' behaviour, not only are we not prevented from reckoning them in, we are obliged to do so. Moreover, our description of the act can encompass many more kinds—and more detailed kinds—of facts about the social context in which an act is performed.

One can of course err (from a utilitarian point of view) in his use of the generalization test, and I am inclined to say that Broad was objecting to such commonplace applications of it. But the possibility of misuse is no argument against the principles involved. If all relevant factors are considered, no false account of circumstances may be supposed. In this context, then, Broad's argument may be deflated from a criticism of the test itself into corrective advice about how to employ it.

Now it might possibly have appeared to Broad (and perhaps to others) that there is a deeper problem for the generalization test, one which would suggest how a 'false account of the circumstances' is entailed. It may appear that adding predicates like 'performed when the practice of A is not general' is logically unacceptable, a paradoxical condition. Perhaps it seems as if the test we employ, in taking others' behaviour into account in the way I have proposed, is 'What would happen if everyone did A when the practice of A is not general?' Or to put it more sharply: 'What would happen if the practice of A were general when in fact the practice of A is not general?'—'What would happen if everyone did A when in fact not everyone does A?'

These certainly sound paradoxical—the suppositions seem self-contradictory. But the foregoing questions are not accurate renderings of the test I propose to make, which is 'What would happen if everyone did $A\bar{C}$?' That is, 'What would happen if everyone *who had occasion to do so* did A *under these circumstances*, when not everyone is doing A (when the practice of A is not general)?' The supposition is, not that everyone will do A, nor that everyone who has occasion to do A will do A, but rather that everyone who has occasion to do $A\bar{C}$ will do $A\bar{C}$. The condition, $\bar{C}$, that A is not generally performed, rules out the possibility that everyone who has occasion to do A will have occasion to do $A\bar{C}$. This is precisely the sort of condition we require—and there is nothing paradoxical about it.

The sense and naturalness of such a test has already been suggested in examples, but should be emphasized once more. Suppose that A stands for 'keeping promises'. In adding the condition $\bar{C}$ we are distinguishing between (1) AC: keeping promises on those occasions when others are, generally, keeping promises (i.e. when the practice of promise-keeping is sufficiently dense to produce the associated threshold effects), and (2) $A\bar{C}$: keeping promises on those occasions when others are *not* generally keeping promises (i.e. when the practice of promise-keeping is not sufficiently dense to produce the associated threshold effects). The condition is a plausible one. Moreover, it is crucial in dealing with acts that have associated thresholds; and it is one that must be considered therefore when a form of utilitarian generalization is properly applied.

Many writers do not consider *facts* about others' behaviour as we have, but consider *beliefs* instead. Harrison, for example, deals with others' behaviour only in this way, acknowledging that such considerations do not constitute extensions to, but rather implications of, the appropriate criteria of relevance. He argues that

> the probability or otherwise of other people doing what I do does have a bearing on my duty to do an action (or to refrain from doing it) if it would have good (or bad) consequences if everybody else did the same. It is true that, if I only have good reason for thinking that other people will not do what I do, then my duty to be just [i.e. to follow the rules *in vacuo*] in a hard case [with negative simple utility] still applies. For other people's reasons for thinking that theirs will not be the general practice are as good as mine and, if everybody failed to apply a rule in a hard case merely because they had good reasons for thinking that others would not do the same, bad consequences would result. But if I had conclusive reasons for thinking that other people would not do the same, then it would be my duty to relieve the hard case. For only one person can have conclusive reasons for thinking that others will not relieve the hard cases he relieves, and, from one person's relieving hard cases, no disastrous consequences follow. (p. 128.)

In other words, one's beliefs about others' behaviour are to some

degree relevant. However, on strictly general utilitarian grounds Harrison argues that in some cases, whereas 'knowing that others will not do *A*' is sufficient to justify one's doing *A*, merely 'having good reason to believe that others will not do *A*' is not. For the latter is compatible with others actually doing *A*, whereas the former is not. Harrison's formulations do not take thresholds into account, and therefore certain qualifications are necessary; but his direction seems clear and reasonable.

Nevertheless there are two ways of viewing Harrison's position—ways which he fails to distinguish. The start of the passage suggests that 'having good (or conclusive) reason to believe *p*' is a kind of modal operator, a way of dealing with probabilities. Viewed in this way his argument becomes: if one knows that *p*, then necessarily *p* (it follows that *p*), and there is no risk in acting accordingly—i.e. there will be no general practice of *A* and hence no bad consequences; but if one merely has good reason to believe that *p*, then probably *p*, and there *is risk* in acting on the assumption that *p*. Probabilities are dealt with in terms of *good reason for thinking*.

Note, however, that Harrison's argument is incomplete in the latter case. The inference to be drawn, presumably, is that one is not justified in doing *A*, in taking the risk, since if everyone did *A* under such circumstances bad consequences would result. Now this would be sufficient to render this performance of *A* wrong only if the principle applied is probabilistic, that is, only if a sufficient condition of an act's being wrong is that it probably has a negative generalized utility. For even if one takes the risk, the actual circumstances may be such that others don't; thus, unless the probability qualification is added, we could proceed to specify the circumstances further. Indeed even in this case, where a probability qualification is assumed as part of the principle, we would be obliged to ask whether (or to what degree) there is probably such a risk. For if it is merely that others could (possibly) do the same, but most likely will not, then the act may be justifiable.

(It is also possible that Harrison is dealing, not with the wrongness of acts, but with questions of praise and blame. That is, he may have in mind that, if one knows that *p*, then one is justified in, cannot reasonably be blamed for, doing *A*; and that,

if one merely has good reason for believing that *p*, then one is not justified in doing *A*, i.e. one can be blamed for doing *A* since it means taking too much of a risk—even though doing *A* in the actual circumstances might not be wrong.)

Harrison's argument may be viewed in another way. In the foregoing interpretation I took certain locutions—'knowing that', i.e. 'having conclusive reason for believing that', and 'having good reason for believing that'—to be modal qualifiers of descriptions of circumstances in which the act is performed, circumstances related to others' behaviour. But Harrison also seems to use these locutions as parts of general utilitarian specifications of the agent, that is, as conditions of his beliefs. He might be understood as speaking of doing *A because* one knows (or believes) that the practice of *A* is not general. And it is possible that such conditions can, very broadly speaking, be understood as causes of acting in a certain way. ('If he had not believed that *p*, he would not have done *A*.' 'When he realized that *p*, he was at once determined to go ahead and do *A*.') This interpretation is far less plausible than the first, but my point is simply that the two must be distinguished. In general, some disutility will flow from acts as a result of certain properties or states of the agent. These should be distinguished from disutilities resulting from the social context in which an act of a certain kind is performed—a social context which is largely independent of the state of one's beliefs, one's motives, and so on.

If one knows that *p* then it follows that *p*. But the reverse does not hold. This shows that in considering facts about others' behaviour, the actual social context, I have employed different, non-equivalent predicates, and that my descriptions are in this respect more inclusive. Perhaps it will be objected, by those who wish to use belief-specifications, that my descriptions heretofore have been too inclusive, that we can in practice only be guided by the beliefs we possess or by our (more or less) good reasons for holding them. I would insist none the less that we allow specifications of the actual states of affairs, admitting that mistakes can occur. This approach allows us to deal with every condition included by any other approach (beliefs are in fact held; probabilities can be a measure of the possibility of error), and it allows us to entertain a notion of the full import of these principles with

no information deficit. We should want to preserve an approach that allows us to explain how one can do the wrong thing, for example, on the basis of mistaken beliefs or on the basis of good but misleading reasons for believing.

Some have argued, on the other hand—in a way designed, it seems, to exclude consideration of others' behaviour generally—that beliefs about others' behaviour as part of the circumstances of an act cannot be considered at all. We shall conclude our discussion by examining two such arguments.

Singer offers one argument (pp. 147–50) in response to Sidgwick. Sidgwick had claimed that a presumption against acts *A* (based upon the undesirable tendency of *A*) could be rebutted if three conditions obtained. First, the instance of *A* in question must have the best simple utility of the alternatives; secondly, the agent must know that his maxim to do *A* is not universally accepted —that is, his maxim is to do *A* under such circumstances, which precludes that a maxim to do *A* is universally accepted; and thirdly, the agent must have a 'reasoned conviction' that his doing *A* will not significantly influence others to do *A*. Singer objects that

> this clearly will not do. For these conditions will apply in nearly every case, and hence one could justify practically anything on this basis. Everyone could argue in the same way, for everyone can have good reason to believe that the act he is contemplating will remain exceptional, that his doing it will not tend to make it widespread. (p. 148.)

This sounds much like Harrison when the latter rejects 'good reasons for believing' in favour of 'knowing'. But Singer's argument is not merely that Sidgwick's three conditions are too weak; it is not merely that everyone might act upon such reasoning and do *A* with bad consequences. For Singer also says that it does not make any difference

> whether one knows for certain, or merely has a reasonable belief, that his action will remain exceptional. No act can be justified simply on the ground that it will not tend to become widespread. . . . For if this reasoning would justify any one such action, it would justify every such action, and this is self-contradictory. (p. 148.)

Wherein lies the 'self-contradiction' is a mystery that goes

unexplained. At times Singer suggests that he means 'self-defeating'—and this suggests that bad consequences would flow from everyone's accepting some spuriously utilitarian chain of reasoning. (See, e.g., pp. 22, 87, 89, and esp. 258.) But in any event, it obviously makes a great deal of difference whether one knows or merely believes that others will not do the same; for if one knows then it follows that others will not do the same. The act in question is then described differently. It is not—as Singer suggests—that every A is justified by this kind of reasoning, but rather that every $A\bar{C}$ is justified. The condition that one knows that the practice of A is not (and will not be) general precludes that it become general; there is no question here of justifying A. There may be something unfair or otherwise wrong about this kind of reasoning (or about action on the basis of it), but that is quite another matter. The point is that the reasoning suggested by Sidgwick can be rendered in strictly utilitarian terms—indeed, in strictly general utilitarian terms.

Singer does not rest his case there, however. He also appeals to an argument offered by A. K. Stout which is directed (fittingly enough) at Harrison. Stout claims:

> Universalising the agent's knowledge or ignorance about what others will do involves a vicious circle or, rather, 'shuttle'. If part of *my* special circumstances is my reasoned belief about how others will act in the same circumstances, then part of *their* special circumstances is their belief about how others, including myself, will act. In judging how they will act in their circumstances I must, then, judge how they believe I will act; and they in their turn must have a belief about what I believe they believe. And so on.
>
> Another way of putting this logical difficulty is that as soon as you make the agent's belief about how others will act a part of the special circumstances, you cannot properly speak of others as in 'the same' or exactly similar circumstances. For their belief includes a belief about me, and my belief does not include a belief about myself. ('But Suppose Everyone Did the Same', *Australasian Journal of Philosophy*, xxxii (May 1954), 18–19.)

This argument rests upon several related confusions. Stout

argues that a certain kind of circumstance cannot be 'universalised'—that is, that this circumstance cannot be specified in such a way as to avoid some logical absurdity. One difficulty turns upon our understanding of three expressions: 'my', 'others', and 'the same circumstances'. A second difficulty concerns the relation of beliefs to grounds of beliefs.

Consider the first question, whether that part of *my* special circumstances, which consists in *my* reasoned belief about how *others* will act in *the same circumstances*, is 'universalisable'. The first point to be noted is that 'the same circumstances' must be taken to signify the same circumstances *in vacuo*—that is, all the relevant circumstances surrounding the act in question except for those special threshold-related circumstances of others doing or not doing an otherwise similar kind of thing. In other words, at this stage our reference to the same circumstances involves absolutely no reference to beliefs (or even facts) about others' pattern of behaviour in the relevant respects.

'Universalisability' in Stout's strictly logical sense depends, then, upon whether the inclusion of the special, i.e. threshold-related circumstance involves adding to the description *in vacuo* anything logically vicious or non-universalizable. And it seems clear that it does not, in view of what might be called the effective extensions of 'my' and 'others'. These systematically vary from case to case. In order to be able to universalize the special circumstance we must be able to give a sense to it for each agent who will have occasion to do the act described *in vacuo*. But there is no difficulty in this. In relation to me, the 'my' refers to my own reasoned beliefs (i.e. the agent's), while everyone else, including Jones and Smith and Brown, is covered by 'others'. In Jones's case 'my' is used for Jones (and not for me) since Jones is then supposed to be the agent, or potential agent; and 'others' include Smith and Brown and me. And so on.

To suppose that there is anything vicious (non-universalizable) about a specification of this kind is tantamount to supposing that there is something essentially vicious and misleading in, e.g., each person's first-person use of 'I' for himself alone. The 'my' does not apply to the same person's beliefs in each case—and there is no difficulty in this; indeed, that is just the function of such an expression. Similarly, the effective extension of 'others'

varies from agent to agent, in each case it includes and excludes a different class of individuals. It should also not be supposed that 'others' has the force of 'everyone' when the latter term is used to include the agent. Perhaps it will be helpful to suggest finally, on this point, an aseptic characterization of the special circumstance: 'the agent's reasoned belief about the actions of those other than the agent who will have occasion to do the sort of thing in question'.

The second difficulty suggested by Stout concerns the expression 'the same circumstances' when that is used, not for a description *in vacuo*, but rather for the whole relevant description including the special circumstance in question. It may be granted that others' beliefs can include a belief about me while my belief cannot include a belief about myself (in this connexion). But it cannot be granted that by universalizing the agent's special circumstances and adding a description of them to the description *in vacuo* one gets oneself tied up in such beliefs about oneself. To suppose that one does get so tied up is to confuse 'believing p' with 'one's grounds for believing p'.

Stout seems to feel that one's grounds for believing that others will act in certain ways necessarily include beliefs about their beliefs. But this is not so. One may have, and often has, many types of grounds for beliefs about others' behaviour; for example, observation, experience, statistics, social theory; and one's grounds need never include beliefs about their beliefs. Nor need their beliefs refer back to one's belief. So at best, Stout's argument would hold only against a severely restricted class of cases. But it need not hold at all. For when one specifies that the agent believes that others are (or are not) doing A, he makes no reference to others' beliefs; he refers solely to their actions. To suppose that such a reference to others' beliefs must be made is to suppose that it is necessarily implicit; it is to suppose, perhaps, both that one's grounds are always only beliefs about others' beliefs and that one's beliefs are equivalent to their grounds.

There are, of course, certain problems in determining the actual social context, how others are or will be behaving, whether or not they are doing or will do A. We may have insufficient knowledge. But this is compatible with others characteristically acting in certain determinate ways. The patterns of behaviour

may not be fixed; habits may be in flux. This is the sort of special case regarding which Stout's argument suggests interesting questions: others may not have decided what to do, and part of their grounds for deciding may be the commitments, even the beliefs, of others. But these cases raise no difficulty in principle for treating others' behaviour (1) factually, in respect of how they actually behave, (2) probabilistically, perhaps using belief-specifications modally, (3) as internalized in the agent's situations, in respect of his beliefs about the social context.

D. *Extensional Equivalence*

The argument thus far has these consequences:

(1) If actions are viewed as completely relevantly described, the generalized/simple utility relation is linear. That is to say, non-linearity results from a failure to take all relevant factors into account; when it seems that the G/S relation is non-linear, then we know that some aspect of the situation has been overlooked. Accordingly, acts have (*a*) desirable effects if, and only if, they have desirable tendencies; (*b*) undesirable effects if, and only if, they have undesirable tendencies; and (*c*) indifferent effects if, and only if, they have indifferent tendencies. For linearity entails that there is no qualitative difference between the generalized and the simple utilities of given acts.

(2) All analogous non-comparative principles are extensionally equivalent. For when such principles are applied, only the quality of the utility of an act is taken as relevant to its rightness or wrongness.

(3) But what of comparative principles? Here the situation is more complex. Notice first, however, that the traditional case against simple utilitarianism rests on the contention that there are certain acts that we ought to perform (all things considered) which do not have good consequences, and that there are certain acts that we ought to refrain from performing (all things considered) which do not have bad consequences. Harrison, for example, accepted this as true, and tried to account for it by reference to the desirability or undesirability of tendencies instead of effects. Thus he claimed that some acts without bad consequences have undesirable tendencies and that some acts without good consequences have desirable tendencies. But the

argument for utilitarian linearity shows that there are no such acts (all relevant factors considered). Thus one kind of argument for general *vis-à-vis* simple utilitarianism simply will not work.

But comparative principles deal ultimately, not with the quality of utilities, but rather with the relative utilities of acts. Here is where the complication arises. I want now to suggest a faulty continuation of the argument for equivalence, to show the flaw in it, and then to complete the argument in a more adequate way.

Suppose that we seek to determine the best course of action among a set of alternatives. This task is appropriate when we are applying (AU) and (GU), which assess acts on the basis of their relative simple and generalized utilities respectively. (Similar considerations apply to other comparative principles.) We have a set of alternatives $a_1, a_2, \ldots, a_m$. Suppose that the simple utilities of these acts are unequal and that therefore they can be ordered decreasingly, $\mathsf{S}_1, \mathsf{S}_2, \ldots, \mathsf{S}_m$ (such that S_{i+1} is less than S_i for any i from 1 to $m-1$). Now for there to be equivalence we must be able to list the generalized utilities of these acts in the same decreasing order: $\mathsf{G}_1, \mathsf{G}_2, \ldots, \mathsf{G}_m$.

Will the ordering necessarily be the same? We might suppose that the ordering will be the same since in *each* case we know (from our argument for linearity) that $\mathsf{G}_i = n \times \mathsf{S}_i$. But to generalize from this in answering the present question about ordering would involve a serious mistake. For here, corresponding to the subscripts '1', '2',..., 'm', we have acts of different kinds. The subscripts represent different teleologically homogeneous classes of acts. n is the number of acts in a given teleologically homogeneous class. But n is not a constant: the value of n depends upon various contingent factors, e.g. the particular thresholds, the number of occasions for doing this or that sort of thing, and so on. Not all these classes need be of the same size. We certainly cannot normally assume that they are of the same size. They might be unit classes and then of the same size; but we can no more assume that n equals 1 than that it does not.

Linearity characterizes the relation between the G and the S of a given act such that $\mathsf{G}_i = n_i \times \mathsf{S}_i$. But since n_1 need not be the same as n_2 (and so on), it seems that the S-ordering can differ

from the G-ordering. For while S_2 is less than S_1, G_2 might be greater than G_1. Hence it may seem that there can be this range of non-equivalence: that, while the S and G for a given act must be of the same quality, the grading of acts as more or less useful (and thus as more or less right or wrong) allows for differences between analogous comparative principles.

But consider what such a difference would entail. If a_1 is selected from the m acts as the best course of action by the application of (AU), the simple utility of that act is greater than that of any alternative. Suppose now that G_2 is greater than G_1 and thus that a_2 is selected as the best among the alternatives by (GU). But S_2 must be less than S_1. Hence if (GU) were followed correctly in this and every other relevantly similar case, the total consequences produced would be $n_2 \times S_2$. Thus, even though S_2 is less than S_1, if (GU) were followed when acts like a_1 are alternatives to acts like a_2, S_2 would be produced instead of S_1. Hence if everyone follows (GU) and does acts like a_2 instead of acts like a_1, the best *possible* results will not be produced, since a_1 acts could have been performed. This is contrary to the requirements of (GU). Something has gone wrong.

The solution, I think, is this: although the value of n can vary, it does not follow that in such cases, among a given set of alternatives, it is not a constant. I would argue that it is a constant, in effect, in such cases. Notice that each of the acts, a_1, a_2,..., a_m, has a complete general utilitarian description 'A_1', 'A_2',..., 'A_m'. But it is an essential part of the circumstances surrounding a_1, for example, that acts like a_2,..., a_m are open. The causally and teleologically significant features of the circumstances surrounding a_1 would be different if (some of) these alternatives were not open and others were. In other words, included (or implicit) in the description 'A_1' of a_1 is a reference to the *kinds* of alternatives open. The descriptions, 'A_1', 'A_2',..., 'A_m', are *internally related*. From this it follows that, for any given set of m alternatives, the value of n for the several kinds of acts is the same when the acts are completely relevantly described.

This reasoning is quite general and implies that the ordering of simple and generalized utilities is the same. I conclude therefore that analogous principles of these two kinds, whether comparative or non-comparative, are extensionally equivalent.

Hence there is no substantive choice between these two kinds of utilitarianism.

The conclusion amounts to this: if we consider all the relevant facts (and utilities) in any given case, it matters not whether we ask 'What would happen if everyone did the same?' instead of 'What would happen if this act were performed?' It matters not, if the considerations employed are assumed to be strictly utilitarian.

IV

RULE-UTILITARIANISM

WE shall now talk of moral rules. In the first place, most proponents of utilitarian generalization have viewed such a principle as the ground of moral rules or else have been concerned to account for the strength of certain purported obligations which (it is argued) cannot be accounted for by Act-Utilitarianism. These obligations may be expressed in general judgements about the rightness or wrongness of acts—judgements which can be justified by appeal to utilitarian generalization and which can plausibly be construed as moral rules. Part of my objective in this chapter is to show how such judgements can be generated.

Connected with this topic is the relation of utilitarian generalization to rule-utilitarianism. I shall discuss two main types of rule-utilitarianism: one, which I shall call *primitive* rule-utilitarianism, is simply utilitarian generalization applied in a certain way; another, typified by 'Ideal Rule-Utilitarianism', will be contrasted with primitive rule-utilitarianism, and thereby with Act-Utilitarianism.

In order to discuss the various types of rule-utilitarianism, it is necessary to distinguish different ways in which rules can be 'grounded in' utility. I shall distinguish, for example, between *theoretical* (or theory-dependent) rules and merely *cautionary* rules (rules of thumb). We may say, for the present, that theoretical rules are rigorously justified by a given principle; they are normally incorporated essentially into a rule-utilitarianism properly so-called; they are rules on the basis of which the rightness or wrongness of particular acts is determined and they often cannot be dispensed with in the application of such a theory. Cautionary rules, on the other hand, are practical aids that are recommended or adopted for the purpose of most efficiently and correctly applying and acting upon a given moral principle when the principle could theoretically be applied directly to determine the rightness or wrongness of acts. As we shall see,

for example, the theoretical rules of Ideal Rule-Utilitarianism cannot be dispensed with in applying that theory, for on such a theory the rightness or wrongness of acts is not dependent upon some utilitarian characteristic of acts themselves (such as their simple or generalized utilities), but is rather dependent upon some utilitarian characteristic of the set of rules under which the acts fall (its *acceptance-utility*).

In the case of Act-Utilitarianism, by contrast, rules can theoretically be dispensed with. For on such a theory the rightness or wrongness of an act depends on some utilitarian characteristic of the act itself—its relative simple utility—and no reference to rules is essential. But there are none the less at least two ways in which rules can be associated with Act-Utilitarianism and these must be separated for the sake of clarity. Firstly, we can have theoretical Act-Utilitarian rules—rules that can be rigorously justified by reference to the Act-Utilitarian principle and that indicate exactly the implications of the principle for particular acts. A theoretical Act-Utilitarian rule against lying would have a form like this: 'Lying is wrong, except when its simple utility in a particular case is greater than the simple utility of veracity, all things considered.' This obviously is a trivial kind of rule since in each case the rightness or wrongness of the given act depends upon its own relative simple utility. Secondly, there are the cautionary rules which are more commonly associated with Act-Utilitarianism—rules that are taken as summaries of what generally or usually has been the case, rules that indicate that acts of certain kinds are normally right or wrong, that in the past they have usually had (relatively) good or bad effects, and that they therefore may be expected to be right or wrong in the future. Such rules may be employed as practical aids in deliberation and criticism, but acts are not definitively right or wrong according to such rules so far as Act-Utilitarianism is concerned. Because cautionary rules are general and presumably based upon experience, they may be called 'generalizations'. But such a label should not suggest that these rules are based upon generalized utilities or anything of that kind. Moreover, such rules cannot be rigorously generated from the Act-Utilitarian principle.

Our main concern will be theoretical rules: first, those that can be generated from (or justified by reference to) utilitarian

generalization; and second, those incorporated into the more prominent forms of rule-utilitarianism. Rules of the first kind are based, not on the simple utilities of acts, but rather upon the effects of a number of acts which might be performed—upon their generalized utilities. These are *primitive* rules. Whole sets of such rules justified by a given form of utilitarian generalization yield a primitive rule-utilitarianism. By comparing these rules with the rules of other types of rule-utilitarianism we can determine the characteristic differences between the latter, on the one hand, and simple and general utilitarianism on the other.

The fact that a type of rule-utilitarianism can be developed from utilitarian generalization allows us to say that nothing is essentially lost or gained—nothing is changed—by appealing to rules as such instead of applying the test of utility directly to acts. The forms of primitive rule-utilitarianism imply exactly what is implied by the forms of simple utilitarianism. But we shall also find that other types of rule-utilitarianism are not extensionally equivalent to primitive rule-utilitarianism and therefore are not extensionally equivalent to simple utilitarianism. A substantive alternative to Act-Utilitarianism on utilitarian grounds is therefore possible. But the differences between these types of theory are not obviously favourable to rule-utilitarianism. This fact will lead me to inquire how a utilitarian can reasonably and consistently embrace a kind of rule-utilitarianism that varies significantly from simple utilitarianism. I shall attempt to reconstruct a mode of argument that seems to be used by utilitarians against Act-Utilitarianism and in favour of alternative theories—a mode of argument both the premisses and conclusions of which do not seem to be tenable.

A. *Primitive Rule-Utilitarianism*

It would seem that we are immediately blocked in our efforts to transform utilitarian generalization into a kind of rule-utilitarianism. For these principles apparently imply nothing whatever about moral rules; rules are not mentioned, no reference is made to them in the formulations we have given the principles. Nevertheless, from our analysis of these principles it is easy to see how a certain mode of application of them would yield judgements which may plausibly be called moral rules.

One type of rule which we often think of when we speak of rules, moral or otherwise, has the form 'Do *A*' (or 'Don't do *B*'). Thus we can think of rules such as 'Tell the truth', 'Don't break promises', and so on. Such rules could of course be qualified in a number of ways; exceptions could explicitly be indicated, or the grounds upon which exceptions might be determined could be mentioned; or most simply we could add 'other things being equal' or an equivalent. Rules of this rather loosely defined form may be called *prescriptive* (or injunctive). This is a prominent kind of rule; but it is not the only kind.

When we speak of rules we also sometimes think of what I shall call *ascriptive* rules, such as 'Lying is wrong.' These have the basic form '*A* is *F*' (or '*A* is not *F*'), which is subject to various sorts of qualifications. This is the kind of rule with which we shall work; it is the kind most suited to the explication of a primitive rule-utilitarianism from utilitarian generalization, largely because I have formulated the latter principles in terms of 'wrong'. In concentrating upon this form of rule, I am not suggesting that this is the only or the most interesting or the most important form of rule or of moral rule in particular; I am not suggesting, for example, that the ascriptive form is a paradigm form for moral rules, nor that it is essential in an adequate analysis of moral reasoning. Nor am I suggesting that other forms of moral rules cannot be employed in the construction of a rule-utilitarian theory, or in the development of one from utilitarian generalization. The kinds of rules which can be employed depend largely upon our formulation of the basic principles.

We may say in general that there is a presumption raised against any act which satisfies a general utilitarian description (not necessarily complete) with which is associated an undesirable tendency. It is important now that we allow ourselves to entertain descriptions which are relevantly *in*complete, for then the way is open to rebut such a presumption—and this is an important part of the present programme. Suppose that we are interested, for example, in certain given kinds of acts such as lying, the breaking of promises, black-marketeering, walking across lawns, and so on. We can use such descriptions as starting-points in the construction of rules. (Such a set may very likely prove to be inexhaustive, so we shall have to add to this collection of starting-

descriptions in constructing a set of rules.) I shall limit my attention to acts which may be presumed wrong.

We begin with such given, usually very general, descriptions and see if they are relevant and whether presumptions can be established against such acts. The presumptions can then be expressed in the main or ascriptive clauses of moral rules, and these rules will be qualified so as to incorporate the equivalent of rebuttals.

Take the case of lying. If the tendency of lying is undesirable then we can presume it wrong. But lying is not necessarily—indeed not always—wrong. We cannot consistently generate simple rules of the form 'Lying is wrong'—without qualification. For 'lying' is a description that is true of many acts, among them acts the complete relevant descriptions of which have tendencies that are not undesirable (or not worse than those of the alternatives, and so on—depending upon the principle being applied). Thus, given our strong sense of 'wrong', we must qualify the ascriptive rule against lying.

Some qualifying clause is essential whenever such a judgement-rule is justified by appeal to a form of utilitarian generalization because only a presumption has been established. For some acts to which the description 'lying' applies, the presumption against them will not be conclusive; for some acts this description is significantly incomplete.

What kind of qualification? Once we speak of presumptions which can be rebutted, the most obvious candidate is a *ceteris paribus* condition which is added to and thereby qualifies the main, ascriptive, clause of the rule. This would yield 'If other things are equal, then lying is wrong.' But two things must be noted about rules qualified in this way. First, we shall be concerned only with that function of the *ceteris paribus* condition that is necessitated by the incompleteness of the starting-description. Thus, a *ceteris paribus* condition is required in those rules of this kind that are justified by strong as well as weak principles. But in the case of strong, as opposed to weak, principles the *ceteris paribus* conditions could be dispensed with if the descriptions were in effect made complete, by more precisely qualifying the main clause in ways which I shall presently suggest. Rules derived from weak principles, however, would

require a *ceteris paribus* condition in any case, since that condition would be transmitted to the general judgement-rule from the weak grounding principle itself, just as it is transmitted to particular judgements. In the following discussion I shall ignore this second aspect or function of the *ceteris paribus* condition, for I wish to suggest how the rules can be so qualified that a *ceteris paribus* clause of the first kind is not required. In effect, then, I shall be explicating only rules that are justifiable by strong principles. The rules that are justifiable by weak principles are not far to seek. To determine them, we need merely add *ceteris paribus* conditions to rules that are justifiable by the corresponding strong principles.

Secondly, the *ceteris paribus* clause is neither the only nor the most illuminating type of qualification that can be added. I shall show how such a condition can be displaced by more explicitly determinate ones.

While on the topic of *ceteris paribus* rules it is worth noting that one interpretation of this condition is inappropriate to rules justified by appeal to utilitarian generalization. It has been suggested by Singer (pp. 99–100) that moral rules have the form '*A* is generally (or usually, or probably, &c.) wrong (or right).' Now this is a common enough form of moral rule; it seems suitable, for example, for cautionary or summary rules. But such terms as 'usually', 'probably', and 'generally' are not appropriate in the present context. First, it should be reaffirmed that 'other things being equal' is merely one possible formulation of the *ceteris paribus* condition. We might alternatively employ notions such as 'prima facie . . . (e.g. wrong)', 'may be presumed . . . (e.g. wrong)', 'presumption . . . (e.g. against)', 'good reason . . . (e.g. against)', and so on. For our purposes, nothing hangs upon the choice within this range. But when presumptions are established in the way I have suggested, on the basis of the value of the tendency of a generic description in connexion with a form of utilitarian generalization, we cannot suppose that such acts are generally (or usually or probably) wrong. A presumption against lying, for example, is fully compatible with most instances of lying, or lying in most kinds of circumstances, not being wrong. We might say that presumptions as such have no quantitative implications. Our understanding of the *ceteris paribus* qualifier

must be in quantitatively neutral terms, such as is provided by the 'other things being equal' formulation. Conversely, if we want to explicate rules of the form '*A* is generally wrong', we cannot do so by means of any straightforward application of utilitarian generalization.

It would be misleading to leave the rules at this stage of analysis. They can be explicated further, two means of which I shall now indicate.

Concise rules first. From our analysis of utilitarian generalization we know, not only how presumptions can be established, but also how they can be rebutted. From the specific dimensions of a given principle we can determine the exact kind of grounds which can be employed in rebutting such a presumption. Such grounds, which I shall call *exception-making criteria*, can be appended to the ascriptive clause (in place of a *ceteris paribus* condition) to yield a concise (but theoretically complete) rule.

Consider for example the simplest kind of case, the rule against lying that presumably is justifiable by a (strong) negative non-comparative principle. The concise form of such a rule is 'Lying is wrong, except when its generalized utility in a given case (in respect of a complete general utilitarian description) is not negative.' This rule tells us in schematic terms but none the less explicitly all that is implied about lying by this form of utilitarian generalization. The new qualification tells us what kinds of considerations are relevant and sufficient for our purposes.

Notice how such a rule is generated. We must first ascertain (we must know or be justified in assuming) that lying has a negative generalized utility. Consequently, whether such a rule against lying can actually be generated from a form of utilitarian generalization depends upon matters of fact and our evaluation of effects. Similarly, it remains to be seen what exceptions to such a rule (i.e. qualifications upon the ascriptive clause or rebuttals of the presumption) could be justified by a particular form of utilitarian generalization. In general all we can say is that the exceptions will accord with those cases that would be found not to be wrong by a direct application of utilitarian generalization.

This type of rule suggests a reformulation of utilitarian generalization into primitive rule-utilitarianism. We can view the

principles as schemata for the generation of such primitive rules, along the following lines: 'If the tendency associated with a given description *A* is undesirable, then a rule of the form "*A* is wrong, except . . ." can be justified.' The appropriate exception-making criteria must be filled in, according as the principle in question is negative or positive, comparative or non-comparative.

(Actually, this schema is insufficiently general: it fails for example to exhaust the implications of the positive and comparative forms, for these types of principles allow presumptions against acts which have indifferent or even desirable tendencies. We can therefore do one of the following things. Taking a cue from Harrison, in his rough draft of a form of utilitarian generalization (p. 114), we might have three rule-schemata corresponding to three kinds of ascriptive rules. The main clauses of such rules would express presumptions for, against, or indifferent with respect to acts having some generic descriptions which are associated with desirable, undesirable, or indifferent tendencies, respectively. Thus, we could have '. . . wrong, except . . .', '. . . wrong not to do, except . . .', and '. . . not wrong, except . . .' rules. The exception-making criteria to be appended will vary accordingly. Alternatively, we might retain only the '. . . wrong, except . . .' form and require that presumptions be based, for example, upon relative generalized utilities for the application of such principles as primitive rule-utilitarianisms. This approach has the advantage of economy in the number of kinds of rules, but has the disadvantage of excluding non-relative descriptions such as 'lying'—descriptions in which we are most usually interested. For theoretical purposes the choice is a matter of detail. I shall continue to employ as examples only cases of '. . . wrong, except . . .' rules which are based upon non-relative descriptions.)

I want now to indicate the alternative schema for theoretically complete primitive rules. Concise rules have the virtue of brevity; they tell us all we need to know in a schematic way about a given kind of act. But this virtue has a high price, for sometimes we want to know whether a given set of circumstances constitutes an exceptional condition, i.e. a condition in which a presumption can be rebutted. For the latter purpose we can employ *expanded*

rules. Expanded rules are in effect developed out of concise ones: the implications of a given concise rule are unpacked and become explicit in the corresponding expanded rule.

What I have in mind is this. Instead of mentioning the exception-making criteria in the qualifying clause, we can list all those conditions the satisfaction of which in given cases suffices to rebut the presumption against the act as described in the main clause of the rule. The qualifying clause becomes a complex disjunctive clause, i.e. a disjunctive list of *exempting-conditions*. Thus, instead of appending to the ascriptive clause a concise qualifier like 'except when the generalized utility . . .', we append the expanded qualifier, a clause of the form 'except when *A* is also *B*, or *C*,..., or *K*'. Here *B*, *C*,..., and *K* (which may be simple or complex descriptions) are exempting-conditions.

For example, taking the simple case of a rule justifiable by a negative, non-comparative principle, if *A* is presumed wrong on the basis of its associated undesirable tendency, then *B* is an exempting-condition if the generalized utility of acts *AB* is not negative. '*A*' might be 'lying' and '*B*' could then be 'performed in order to save another's life'.

(Here we find another technical complication which need not detain us. We could require that exempting-conditions be so specified that '*AB*', '*AC*', and so on, represent complete relevant descriptions of acts. But this may be unnecessarily stringent. We might allow '*B*' to be included if, in virtue of the fact that an *A* is also a *B*, the presumption against the act that is based upon its being an *A* is rebutted. In other words, *B* would be a 'right-making' condition sufficient to outweigh the 'wrong-making' condition *A*. But then such a rule might fail to deal with other instances of *A* which are not made right simply because they are also *B*. I am thinking of cases in which there is more than one wrong-making condition, e.g. *D* as well as *A*, and wherein *B* is not sufficient to outweigh *AD*, although *BE* may be sufficient. We then have a choice. We can include more complex exempting-conditions in the rule so that it becomes, in part, '*A* is wrong, except when *A* is also *B*,..., or *BDE*,....' Or we can construct an additional rule beginning with *AD*, one of the exempting conditions of which would be *BE*, but not *B*.)

All the concise rules of a given primitive rule-utilitarianism

can in principle be developed into expanded rules. But it is important to note what kind of claim I am making in saying this. I am not saying that we can in practice complete all the calculations required to expand a concise rule, that we can complete the listing of all the exempting-conditions of all the expanded rules. I am simply proposing a model for viewing the import of a given principle *vis-à-vis* a given kind of act. The notion of an expanded rule is to be used as a device for conceptually cataloguing presumptions and rebuttals. There is perfectly good sense in claiming that such presumptions and rebuttals are implied by a given principle (in a given context). And if there is sense in such a claim, then it may be granted that these rule-schemata can be useful devices for analysing such implications *en masse*.

I am especially interested in two classes of exempting-conditions that would be included in those rules concerning acts that have associated thresholds. These are (1) what I shall call *maximizing-conditions*: designed to maximize the utility of a practice, to make the best of a generally good situation. These conditions are based upon the existence of a (relatively) good-producing practice or the non-existence of a (relatively) bad-producing practice; (2) *minimizing-conditions*: designed to minimize the disutility of a practice, to make the best of a generally bad situation. These are based upon the non-existence of a good-producing practice or the existence of a bad-producing practice. These classes of exceptions do not exhaust all the conditions which would probably be included in the rules, but they have this significance: they are based upon the pattern of others doing *the same*, where 'the same' represents a description *in vacuo*. (If we strike maximizing- and minimizing-conditions from the rules they become rules *in vacuo*.) These exempting-conditions are therefore crucial to the equivalence of primitive rule-utilitarianism and simple utilitarianism.

I shall now try to indicate by means of an example the nature of maximizing- and minimizing-conditions. It is important to be clear about these because they will be central to our extensional analysis of other kinds (non-primitive kinds) of rule-utilitarianism. Suppose that six of us are driving along in our jointly-owned automobile. It is important for each of us—perhaps for all of us jointly—to get to our destination as quickly as possible. But

owing to our limited resources our car is not the newest and freshest on the road. Its paint is worn and it has no tail-fins or shining hub-caps. But its main fault is that it does not run well. It seems malevolently predisposed to stalling at inconvenient places, e.g. upon ascending hills. And since the battery is weak it can be started again only by brute force: by getting it moving and throwing it into gear.

Suppose further that we know this car and its idiosyncrasies well. Now the car stalls—this time, when we stop along a level stretch of road. Moreover, we have only twenty yards or so to push the car before a descent, and thus we know that we could all get it moving again just by pushing it and coasting it down the hill a short way. But we are all tired, having had to do this sort of thing too often. It will be hard work to start the car. None the less, it can be done. Indeed, we also know that five of us could manage the job. (The threshold effect, getting the car moving, &c., requires less than universal performance.) And whether five or six of us do it, the reward in getting the car started and continuing on our way would undoubtedly outweigh the several efforts that would be involved in doing it. (No question but that everyone's pushing would yield good effects on the whole: a good-producing practice.)

But since five could do the job without undue strain, and since (let us suppose) the relief for one who does not push would outweigh the sum of extra efforts required by the other five in pushing in such a case, a maximizing-condition is satisfied. For the practice of pushing the car would have better consequences over-all if one of us did not push. The good threshold effects would be produced anyway, and the relief afforded by a general but not universal practice would enhance the net-balance of good produced. Thus one exception could be made.

But a problem arises even in such a simplified example which does not arise in the more complex cases such as voting where there can be assumed to be a relation of anonymity among the several agents. For in the present case no one is in a position to exploit the maximizing-condition. One of us could not, for example, simply opt out of pushing for the sake of increasing the net-balance of good produced. Jones wants to rest; but so do all of us; why he? It is possible that more harm than good could

come out of one's insisting upon being excused without a special reason beyond that of increasing the net-balance of good. An argument might ensue, tempers could flare, and a generally bad time would result.

Suppose, however, that others are pushing and that one merely feigns pushing. Feigning also takes some effort, but it might sometimes yield an enhanced net-balance of good. By means of the simple expedient of secrecy one could exploit the maximizing-condition. But a natural objection against such behaviour is that it is not fair. Someone would be 'getting away with' working less at the expense of others. Even if the others did not have to push proportionately harder to make up for Jones's failure to push, Jones's act might still be considered unfair in that the benefits would be inequitably distributed.

The issue of fairness is one to which we shall return. Here I wish merely to acknowledge that the obvious unfairness of Jones's demand or behaviour can itself have disutility. But it is not clear that in all such cases in which one acts alone to exploit a maximizing-condition—even if one acts unfairly—there will be a net-loss of utility. Nor is it clear that the notion of fairness always applies to cases involving maximizing-conditions. If one abstains from voting because of inconvenience when in fact enough others will vote to enable, say, the best man to be elected, it is not at all clear that one's act is unfair (though it may be wrong, although not on utilitarian grounds). Nor does it seem that abstaining would result in a net-loss of utility, not only if one person abstains, but if all others in a relevantly similar position do so as well.

The objection might be made that the primitive rule that would allow such exceptions based upon maximizing-conditions is purportedly moral, and thus must be publicizable. The objection would run this way: if there were a general awareness of such an exempting-condition, if the import of the rule including that condition were made public, it would be impossible to take advantage of the exception. Everyone would try to act on the exception and thus no one would do what was required. Or else the obvious unfairness of the act (exploiting the maximizing-condition) would itself have bad effects.

I shall make just two comments on such an objection for the

present. First, it is impossible for everyone covered by a rule to exploit a maximizing-condition. If everyone tried to do so, the good-producing practice would collapse (e.g. the car would not be pushed at all, or no one would vote). It is not enough merely to think that one is acting upon a maximizing-condition. Whether one can possibly do so depends upon the actual behaviour of others. Secondly, the question whether a rule is publicizable must be distinguished from the question whether it is in fact public. It is perfectly consistent and intelligible for a moral theory to be such that its rules include exceptions based upon whether or not others are generally conforming with, are aware of, or accept the rules. If the theory is effectively publicized and perhaps accepted and acted upon, then some of the exceptions which would be allowed under different circumstances will not be allowed in these.

Let us now consider a minimizing-condition, using the same example. Suppose that three of the six of us who are travelling in the car refuse to push when it is stalled. These three feel that they have had all the pushing they can stand and simply refuse to do any more. They refuse to be convinced by argument; they will not see—or act upon—their own and everyone's best interests. Obviously, the silliest thing one could do in such a case, where five must push for any good effect and only three are willing to do so, is push anyway. If one thought that in pushing one acted upon a primitive rule (doing that which would have the best effects if everyone did it), this would be a mistake. A primitive rule covering such a case would include a minimizing-condition indicating in effect that it would not be wrong not to push when the car would not get started anyway (even though it would be wrong to fail to push when the requisite number were willing and able).

Maximizing- and minimizing-conditions should be kept distinct. Although both depend upon others doing or not doing *the same*, they depend upon such factors in different ways and have different utilitarian points. As we shall see, although maximizing-conditions are somewhat problematic—some would be rejected from a non-utilitarian point of view—they are not generally excluded from the rules of non-primitive rule-utilitarianisms. Minimizing-conditions, on the other hand, which

seem eminently sensible, are not clearly included in the rules of such theories. Thus, the non-primitive rule-utilitarianisms differ from primitive rule-utilitarianism in the wrong ways, as I shall argue.

I should mention finally that what I am calling 'minimizing-conditions' seem to correspond with those sometimes called 'states of nature' by proponents of utilitarian generalization. (See, e.g., Harrison, pp. 125–9, and Singer, pp. 152–61.) I find the term 'state of nature' inappropriate, however, first, because some minimizing-conditions are satisfied in circumstances which are far too mild to fit the catastrophic connotations of such a term as 'state of nature', and second, because the circumstances which would satisfy some minimizing-conditions (e.g. general lying, general promise-breaking) are so serious, so disastrous, as to suggest a state of affairs which cannot intelligibly be supposed to obtain along with ordinary human existence.

Now some general comments on primitive rule-utilitarianism. In the first place, such a rule-utilitarianism can in principle be developed. That is to say, general judgements that are suitably and exhaustively qualified and that may be viewed as moral rules can be justified by any given form of utilitarian generalization. I am not suggesting that we can in practice grind out a set of rules completely and exactly exhausting the implications of such a principle. But this is not because such a set of rules could not be extensionally equivalent to the principle; the difficulty is not a result of the transition to rules (i.e. to intermediary general judgements that may be viewed as rules). It is essentially the difficulty we faced before in considering the application of utilitarian generalization directly to acts: the factors are so complex that we could not necessarily run through all of them in practice.

It may be misleading to speak of *a* set of rules extensionally equivalent to one such principle. For one thing, there might be several, even innumerable such sets for a given principle, since there may be several, perhaps innumerable sets of ascriptive clauses (starting-descriptions) that could be employed in the construction of a set of rules completely exhausting the implications of a given principle. Of course we are not in the least concerned with economy in such a set—we are not concerned, for example, with a set containing the smallest number of shortest

possible rules. We are not concerned primarily because we have imposed no special conditions upon the rules of our primitive rule-utilitarianisms. We are not supposing, for example, that these general judgement-rules match any given set of conventionally accepted rules, e.g. the moral code of a given community. These are not rules in the sense of actually being accepted by the members of a given community; they are judgements that might be viewed as rules and that are also implied by given principles.

Primitive rule-utilitarianisms must also be seen in the light of earlier observations about consistency. If the sets of rules preserve extensional equivalence with the grounding principles, there will be as much danger—but of course no more danger—of inconsistencies in the sets of rules as among the judgements derived directly. The comparative principles seem to be internally consistent; therefore the rules based upon them will be mutually consistent. But we would expect that the strong non-comparative principles would generate mutually incompatible rules; these inconsistencies are localized in 'lesser of two evils' and similar situations, when more than one act is counted as wrong not to do, or when all the alternatives are counted as wrong. Again, the weak non-comparative principles will necessarily yield mutually consistent rules: this is the result of retaining the *ceteris paribus* condition which now functions only to eliminate the inconsistencies in the corresponding strong non-comparative principles.

Most important of all, it follows, from the fact that such rule-utilitarian theories can be developed extensionally equivalent to the forms of utilitarian generalization, that an appeal to rules as such makes no effective difference in the import of a utilitarian theory. For it is not a question what steps our reasoning must take, nor what forms our judgements must assume, but simply the content of the rules and the ultimate import of the principles. Therefore, if rule-utilitarianism is to offer a substantive alternative to simple utilitarianism, there must be special conditions imposed upon the rules so that they are not unrestrictedly grounded upon generalized utility and so that they may therefore diverge from the rules of primitive rule-utilitarianism.

As I have already suggested, however, rule-utilitarians do have an alternative—indeed, an indefinite number of alternatives.

These should not be viewed as modifications of primitive rule-utilitarianism, as theories altered and patched up so as to guarantee a measure of non-equivalence with, say, Act-Utilitarianism. For these non-primitive rule-utilitarianisms have actually been developed in many different ways. Nevertheless, it will be helpful in trying to understand the import of the alternative theories to compare them extensionally with primitive rule-utilitarianism. Thus, my explanation of how rules could be generated from the forms of utilitarian generalization had a twofold purpose: first to explicate the notion that utilitarian generalization can be viewed as the ground of moral rules, and second to show that a mere appeal to rules does not necessarily yield non-equivalence with simple utilitarianism. But my elaboration of the details, my brief discussion of maximizing- and minimizing-conditions, was intended to lay the basis for a systematic comparison of Act-Utilitarianism with its non-primitive rule-utilitarian alternatives. This can be done by means of the primitive rules of (GU), the general utilitarian analogue of Act-Utilitarianism. For our purposes we may restrict our attention now to this form of primitive rule-utilitarianism, which I shall call Primitive Rule-Utilitarianism (PRU).

In order to make perfectly clear the sort of comparison I shall make I want to propose that we view the several alternative theories in the following way. It seems reasonable to suppose, first, that as far as action-guidance is concerned, all the rules of most (perhaps all) alternative rule-utilitarian theories can be translated into one limited normative vocabulary such as ours based upon the term 'wrong'. Regardless of whether the rules are normally in the ascriptive mood or some other (e.g. prescriptive), they can all presumably be translated into the former. For example, acts that on other theories *ought not* to be done, all things considered, or that rules like 'Don't do *A*' tell us not to do, are therefore *wrong* in the translation. Acts we *ought* to do, or we are told to do ('Do *A*'), are acts it would be *wrong not* to do. Other acts are simply *right*. If these theories have second-order rules to adjudicate conflicts among incomplete first-order rules, their directions can presumably be translated down into more full-bodied and complete first-order rules of the kind we have been examining.

Moreover, it seems reasonable to suppose that, in many (perhaps all) cases, the specific rules in the sets justified by these alternative theories can be so generated as to maximally overlap with the set of primitive rules. Of course we do not necessarily have just one set of primitive rules for (PRU), since there may very well be an indefinite number of complete sets containing exactly the implications of (GU), but which can be distinguished by the fact that they are generated from different sets of starting-descriptions. This applies to other theories as well: they may very well have several sets of rules that could equally well be justified, implying exactly the same things but formulated in different ways. All we require for our purposes, however, is that there be one set of primitive rules upon which we can fix, which yields (or is based upon) one set of starting-descriptions and ascriptive clauses.

Now if we take two corresponding sets of rules, a complete set of primitive rules on the one hand, and a complete set of rules for an alternative (non-primitive) theory generated as far as possible by beginning with the same ascriptive clauses as are employed in generating the primitive rules, then we may find two kinds of differences between the sets. There may be *rule-gaps*, wherein a rule of one theory based upon a certain starting-description is simply not included in the corresponding set of rules for the alternative theory; and there will be *exception-gaps*, wherein two corresponding rules with identical ascriptive clauses have different sets of exceptions (i.e. different exception-making criteria which yield differences in the expanded lists of exempting-conditions). We can treat the former as a limiting case of the latter and restrict our attention to one type of difference between the corresponding rules of the alternative theories: exception-gaps. By noting general characteristics of such gaps between the primitive rules of (PRU) and those of alternative rule-utilitarianisms we can determine schematically the differences in import between these two kinds of rule-utilitarianism. We shall see that this sort of contrast is most directly related to those threshold considerations upon which the equivalence argument is based and, among these, those to which the arguments from fairness are most relevant.

B. *Non-Primitive Rule-Utilitarianism*

In the following I shall not consider all the possible varieties of rule-utilitarianism. I shall restrict my attention to the predominant type of theory actually proposed, which has been offered as a utilitarian alternative to Act-Utilitarianism. These non-primitive rule-utilitarianisms may be said to be analogous to Act-Utilitarianism and to (GU)—its general utilitarian analogue—or (PRU)—its primitive rule-utilitarian analogue—in the sense that in each case the notion of *maximizing utility* is centrally placed. The rightness or wrongness of acts turns in some way, directly or indirectly, upon their being conducive to the production of the best effects.

The idiom commonly used in the formulation of rule-utilitarian theories is somewhat different from that which we have been using. But the common, central terms such as 'right' and 'wrong' have the same, more or less technical sense: they are contradictories. For example, the theory with which we shall primarily be concerned, sometimes called 'Ideal Rule-Utilitarianism', may roughly be formulated in the following way:

(IRU) An act is right if, and only if, it conforms to a set of rules general acceptance of which would maximize utility.

An act conforms to a set of rules only if the rules do not imply that the act is wrong; or, in the terminology of prescriptive rules, only if they do not indicate that the act should not (ought not to) be done, all things considered. I shall deal with the notion of *acceptance* presently. Suffice it to say for now that a rule is accepted only if one tries to apply and act upon it. (See R. B. Brandt, *Ethical Theory* (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1959), pp. 396–400, and, by the same author, 'Toward a Credible Form of Utilitarianism', in H.-N. Castañeda and G. Nakhnikian, *Morality and the Language of Conduct* (Detroit: Wayne State University Press, 1963), pp. 107–40, for formulations and development of theories of this type.)

For the sake of contrast, I shall suggest how (AU) and (GU) would look in the characteristic form in which rule-utilitarian principles are rendered. Act-Utilitarianism becomes:

(AU) An act is right if, and only if, its total effects are no worse than those of any alternative.

Similarly, its general utilitarian analogue becomes:

(GU) An act is right if, and only if, its completely relevantly determined tendency is no worse than that of any alternative.

Now let us try to recast (GU) into explicitly primitive rule-utilitarian terms. We discover an interesting and illuminating difficulty. Consider the following principle:

(SRU) An act is right if, and only if, it conforms to a set of rules *general conformity* to which would maximize utility.

Brandt calls this a 'specious' rule-utilitarianism ('Toward a Credible Form of Utilitarianism', pp. 119–23) because 'it has identically the same consequences for behaviour as does act-utilitarianism' (pp. 120–1). Brandt evidently is claiming what I would claim by saying that (AU) and (SRU) are extensionally equivalent. But these principles are not equivalent, and that is because (SRU) is not the rule-utilitarian rendering of (GU). I shall try to show this by reference to maximizing- and minimizing-conditions. The differences between (AU) and GU), on the one hand, and (SRU), on the other, can be shown in relation to systematic differences between the primitive rules of (PRU)—which is extensionally equivalent to (AU) and to (GU)—and the 'specious' rules of (SRU).

The difference is briefly this. Whereas minimizing-conditions can and must be included in the primitive rules, their inclusion in the specious rules is indeterminate. Consider, for example, only those primitive rules against acts A which are based primarily upon the undesirable threshold effects which could be produced by the general performance of A. Minimizing-conditions allow for circumstances in which there is in fact a general practice of A (and thus A's bad threshold effects are produced): acts A would be allowed under such circumstances. That is, in so far as the bad effects of acts A are threshold-related, if the threshold is already passed, then further acts A would not be considered wrong. Minimizing-conditions, as I have said, are designed to make the best of such a generally bad situation.

But to say that minimizing-conditions allowing acts A are based upon the general performance of A is to say, in effect,

that they are based upon a general failure to conform to the rule against *A*. For if the rule against *A* were generally conformed to, then too few acts *A* would be performed to allow the production of the undesirable threshold effects. (Some acts *A* might be performed when the rule against *A* is generally conformed to, for some such acts might be allowed by maximizing-conditions. But when *A*'s undesirable threshold effects are produced, it is not possible that everyone, or most people, are conforming to the rule against *A*. It is not possible that the rule is generally conformed to.)

Now the difference between Act-Utilitarianism and (PRU), on the one hand, and (SRU), on the other, turns upon the fact that in the latter case a *general conformity test* is employed in justifying rules. In determining a set of rules for (SRU), we ask what set of rules would have the best consequences if they were generally conformed to. But if we take two alternative sets of rules, one of which does and the other of which does not include minimizing-conditions, but that are identical in every other respect, then the general conformity test will not distinguish between these two sets. For the test is based upon the supposition of general conformity. Under such a supposition, the minimizing-conditions will simply be vacuous. For the minimizing-conditions are satisfied only when there is not general conformity. Thus, no difference will show up between the rules that contain minimizing-conditions and the rules that do not, under the general conformity test. This means that (SRU) may be regarded as indeterminate for this class of cases, that a set of rules without minimizing-conditions is compatible with (SRU). And thus it cannot be claimed that the implications of (SRU) and those of (AU) are identical, for (AU) is by no means indeterminate for this class of cases.

The point can be put in another way. It is misleading to analyse (PRU) in 'conformity'-terms parallel to the terms appropriate to (SRU). (PRU) is designed to cover all cases from the point of view of the agent's act or choice of action. One of the factors which is given full rein in determining the rightness or wrongness of an act under (PRU) is whether others are not, for example, doing *A* when there is a general rule against *A*. If others are generally doing *A* then the best possible consequences on the whole cannot be produced, in the sense that better

consequences would be (or would have been) produced if fewer did *A*. But given these circumstances wherein *A* is generally performed, the agent's act is assessed according to the best possible consequences that can *then* be produced. A principle like (AU) or (PRU) is so designed that, if it were conformed to, or correctly applied and acted upon, in *any* particular set of circumstances, then the best possible consequences that could be produced under those circumstances would be produced. But the general conformity test employed in applying (SRU) does not provide for all such contingencies. It provides only for conditions in which the rules are generally conformed to. Thus, the implications of (AU) and (PRU), on the one hand, and of (SRU), on the other, are identical only when these principles are in fact generally conformed to. And this is a very special case indeed.

How then are we correctly to formulate (PRU)? Perhaps this way:

(PRU) An act is right if, and only if, it conforms to a set of rules *conformity to which in the case in question* would maximize utility.

It is understood, of course, that the case in question is always to be treated as one among a number of similar cases, i.e. completely relevantly determined acts, classified on general utilitarian grounds.

It is clear why Brandt supposed that (AU) and (SRU) are extensionally equivalent. He was aware that others' behaviour is a circumstance relevant to the efficacy of acts and to their descriptions, and that this factor must, in consistency, be taken into account. (See Brandt, 'Toward a Credible Form of Utilitarianism', pp. 121–3.) Moreover, there is no reason for supposing that (SRU) and (PRU) would differ in respect of maximizing-conditions. For, as I have already indicated, general conformity to a rule against *A* is compatible with some people doing *A*, provided that *A* is not generally performed. Brandt failed to distinguish between two radically different kinds of circumstances in which others doing *the same* must be considered: he failed, in effect, to distinguish minimizing- from maximizing-conditions.

A comparison of (PRU), and thereby of (AU), with Ideal Rule-Utilitarianism can be developed along similar lines. In this case, however, the comparison must be rough, for the ideal rules of (IRU) are determined by a *general acceptance test* which involves more factors than those involved in a general conformity test. A word is in order on the relation of rightness and wrongness to the maximization of utility under (IRU).

Ideal Rule-Utilitarianism (or a paradigm principle of this type) is formulated, once more, as follows:

(IRU) An act is right if, and only if, it conforms to a set of rules general acceptance of which would maximize utility.

On the Ideal Rule-Utilitarian approach, acts are not assessed on the basis of their own simple or generalized utilities. The rules themselves are assessed as functions of the utility of their general acceptance—their *acceptance-utilities*, that is, the value of the effects of the general acceptance of the set of rules. Acts derive their rightness or wrongness from what the validated rules have to say about them.

Now in what does such general acceptance consist? Roughly, it would seem that one accepts a given rule or principle when one adopts and avows or at least makes a reasonable effort to apply and act upon it. This entails some measure of success in doing what the rule or principle requires, that is, in doing right acts only. Not perfect conformity: success cannot be expected in every case. For one can mistakenly infer from a rule or principle things that the rule or principle does not imply. Some allowance must therefore be made for mistakes, and also for failures to do what one concludes, in any case, one ought to do. Thus general acceptance differs from general conformity. In the former case a great deal of conformity is presupposed, but supposedly somewhat less than in the latter—although it is not clear how much less. Thus the ideal rules will not necessarily be the same as the rules of (SRU), since differences in conformity can have significant effects upon the utility of a given rule or set of rules. In other words, the conformance-utilities and acceptance-utilities of given sets of rules can differ markedly. And since this is so, the set having the highest conformance-utility will not necessarily be the same as the set having the highest acceptance-utility.

But in any event, if we judged sets of rules on the basis of the utility that could be gotten from their general acceptance, we would find that two sets of rules, that are exactly similar in all respects except that one set includes minimizing-conditions while the other does not, would not be differentiated. For again, since we are postulating general acceptance in the general-acceptance test, the minimizing-conditions would in general be vacuous. Thus general acceptance of either set would, roughly speaking, have the same consequences.

There is a further reason for supposing in the case of (IRU), although perhaps not in the case of (SRU), that general acceptance requires the omission of minimizing-conditions, and that (IRU) is not merely indeterminate for this class of cases. The principle we are examining is called 'Ideal Rule-Utilitarianism' because the rules are supposed to amount to 'ideal rules' (or, as Brandt calls them, 'ideal prescriptions') for the agent's community. They constitute a kind of utilitarian social policy. (See Brandt, 'Toward a Credible Form of Utilitarianism', p. 109 and *passim*.) They are the rules that, supposedly on utilitarian grounds, it would be best if everyone tried to conform to. They are supposedly suitable for general adoption as moral rules in the given community. Thus there is the image of a public moral code, promulgated, codified. And when one begins to think in such quasi-legal terms, it becomes impossible to allow that minimizing-conditions can be included in the ideal rules. For example, when we frame a law we frame it supposing that it will be generally observed; we do not and cannot allow exceptions merely on the ground that it is not generally observed. (Although they may become and may be officially recognized as 'dead letters', laws do not incorporate minimizing-conditions excusing non-conformity simply because they are not generally observed.) Moreover, since the consequences would in fact be better if no minimizing-conditions ever were satisfied, perhaps in drafting such an ideal moral code one would not entertain exempting-conditions which are based upon a generally bad (non-conformist) state of affairs.

Thus, we can say in general that Ideal Rule-Utilitarianism does indeed provide a substantive alternative to Act-Utilitarianism. But it is not clearly the sort of theory with which utilitarians can be satisfied. For to act upon a theory that does not allow (or is

indeterminate with respect to) minimizing-conditions is possibly to act, from a utilitarian point of view, in a self-defeating manner. An argument could be developed along the following lines, similar to a type of argument which we shall examine in greater detail presently. (1) The point of (IRU) is to maximize utility. (2) But there are some cases in which applying and acting upon (IRU) will not yield the best possible consequences, that is, when (IRU) is not generally accepted. (3) Thus (IRU) cannot be an acceptable moral principle, since in some cases (that is, when it is not generally accepted—which is the normal state of affairs) applying and acting upon (IRU) would be self-defeating.

Rule-Utilitarians have been aware of this problem and sensitive to criticisms along these lines. (See, e.g., Brandt, *Ethical Theory*, p. 400.) The consensus seems to be that minimizing-conditions must, in effect, be acknowledged. But to say this is to say one of the following: (1) Exceptions must be allowed to the ideal rules that are not allowed by the governing principle. This is manifest inconsistency. (2) The ideal rules are (let us suppose) indeterminate for the range of cases in question. Thus the governing principle is unable to account for an important range of cases. This makes the Ideal Rule-Utilitarian position incoherent in a special way. For, since two differing sets of rules can be validated by the general acceptance test, that is, rules that do and rules that do not allow certain classes of exceptions, it appears that according to (IRU) some acts are both right and wrong or neither right nor wrong. The former result is untenable; the latter is in discord with the apparent generality of (IRU), the formulation of which suggests that all acts are either right or wrong.

Brandt has given considerable attention in his recent work to the problem of determining a set of rules 'for an imperfect society'. ('Toward a Credible Form of Utilitarianism', pp. 126–30.) He acknowledges, for reasons which overlap with but are somewhat different from mine, that Ideal Rule-Utilitarianism 'savors a bit of the utopian' (p. 127). He proposes, consequently, a modification of the latter theory that incorporates into the test for the set of rules a consideration of the existing (*de facto*) moral beliefs within a given community. The resulting, more 'credible' form of rule-utilitarianism, which I wish to comment upon briefly, is presented as follows:

> The compromise I propose is this: that the test whether an act is right is whether it is compatible with that set of rules which, were it to replace the moral commitments of members of the *actual society* at the time, *except where there are already fairly decided moral convictions*, would maximize utility. (p. 129.)

The resulting principle is:

(CRU) An act is right if and only if it conforms with that learnable set of rules, the recognition of which as morally binding, roughly at the time of the act, by all actual people insofar as these rules are not incompatible with existing fairly decided moral commitments, would maximize intrinsic value. (p. 129.)

It should be observed, however, that the *de facto* morality of a given community can be grotesquely immoral. Thus, the resulting standards gotten by applying (CRU) might be far too low. On the other side we find that one of the strongest features of moral beliefs in societies like ours is a sensitivity to unfairness and injustice. Consequently, the validated set of rules for such a society will be determined, in part, not merely by considerations of utility, but also by considerations of justice and fairness. On such a theory as (CRU), therefore, non-utilitarian factors are introduced indirectly.

These observations suggest that the rule-utilitarian is gradually forced to forsake a pure utilitarianism for a composite theory. In the next chapter I shall deal in greater detail with some of the arguments from justice and fairness that are particularly relevant to rule-utilitarianism and to utilitarian generalization—arguments which tend to show that a pure utilitarian theory, whether or not it involves rules, cannot be adequate. But first I want to examine the more immediate question, whether principles like Ideal Rule-Utilitarianism can reasonably be regarded as utilitarian. To this we shall now turn.

c. *Utility and Rules*

Because non-primitive rule-utilitarianisms differ substantively from Act-Utilitarianism we can reasonably ask a question which might well have been asked before, but the raising of which was made unnecessary by the arguments for extensional equivalence.

The question is: how is it possible for a utilitarian consistently to embrace a theory that is, or is supposed to be, a substantive alternative to Act-Utilitarianism? For whenever a person successfully applies and acts upon (AU) he performs an act that will produce as good consequences as could possibly be produced under the circumstances. Thus, to the extent that, say, (IRU) diverges from (AU), there will be some cases wherein, if (IRU) is successfully applied and acted upon, the best possible consequences that could be produced will not be produced, or at the very least one has restricted one's choice of action by regarding as wrong acts whose effects would not be worse than those of any alternative.

A strictly analogous comparison could be made between non-primitive rule-utilitarianisms and (GU) by speaking of the results of everyone's doing the relevantly similar sort of thing. But the comparison could be developed more sharply in terms of exception-gaps between primitive and non-primitive rules. Wherever there is an exception-gap between the primitive rules of (PRU) and the ideal rules of (IRU), that signals a case in which successful following of the ideal rather than the primitive rule would forsake utility in the above sense. Thus it would seem that adherence by an avowed utilitarian to a non-primitive rule-utilitarianism such as (IRU) is inconsistent and therefore demands an explanation.

I shall now try to reconstruct one particularly relevant kind of argument by means of which utilitarians may—and apparently sometimes do—justify their rejection of Act-Utilitarianism and their adoption of an alternative 'utilitarian' theory. We can begin with a seemingly straightforward argument against Act-Utilitarianism. It is sometimes claimed (and must be admitted) that we hardly ever have all the information required for the application of (AU), that is, enough information to warrant our assurance that what we infer about the rightness or wrongness of acts is actually implied by the principle. Moreover, time presses and we are often forced to act in haste, sometimes without adequate deliberation. We are also careless and prone to error in favour of our own interests. And sometimes we fail to do what we believe to be right—what we believe to be required or allowed by our moral principles. Thus, from the fact that (AU) is applied and

acted upon (allowing a reasonable margin for failure), it does not follow that the best possible consequences are produced. To try is not necessarily to succeed.

What does this argument prove? I want to distinguish two ways in which such considerations may be employed. For the Act-Utilitarian they assume the significance of practical obstacles to the successful application of his theory. For the rule-utilitarian they may take on a radically different significance: they suggest an argument which, if successful, strikes at the very core of Act-Utilitarianism.

Consider the argument in practical terms first. It is admitted that mere acceptance of Act-Utilitarianism does not maximize utility, for the reasons given above. Since it may also be argued that our chances of success in maximizing utility would be enhanced by our application of a simple set of rules (instead of applying the Act-Utilitarian criterion directly), this may be turned into an argument for 'rule-utilitarianism'. That is, if we want to succeed in producing the best consequences—the 'point' of Act-Utilitarianism—then we should not calculate utilities directly, but should instead rely upon and appeal to a simple set of rules.

But interestingly enough, the Act-Utilitarian is apt to accept this argument, to use it for urging that in practice we observe a set of moral rules. He is apt to recommend that we appeal to *cautionary* rules, rules that can be employed as practical aids in determining the best course of action, rules that are more reliable in their action-guidance than we are likely to be in our deliberations.

In order to understand the Act-Utilitarian's position, however, in order to see how misleading it would be to characterize his employment of cautionary rules as 'rule-utilitarian', it is necessary to draw a further distinction among types of rules. This distinction, between *theoretical* (theory-dependent) and *de facto* (conventional) rules, concerns what might be called their status; it is a distinction cutting across those we have already made which turn mainly upon form.

Some rules owe their very existence to the fact that they are acknowledged, observed, avowed, referred to—that individuals can and sometimes do appeal to them in criticizing others' behaviour and in justifying or criticizing their own conduct. We

determine that such rules exist and what they imply (what their content or import is) by seeing how people behave and how they talk and argue about their behaviour. These are *de facto* rules. This class includes, for example, legal rules, traditions, conventions, and mores. It may be difficult in practice to determine which *de facto* rules belong to a given group or society. Legal rules are generally easiest to pin down since they are normally recorded. Unwritten rules are, however, harder to verify. And in the realm of unwritten rules—including conventional or *de facto* morality in general—we face the difficulty of interlocking sub-groups in a particular society, each with its own complex set of rules. It is also true that the content of some of the rules of most groups is hard to determine—or indeterminate to some extent—because of a lack of unanimity among the members of the groups. And some rules receive more lip-service than conscientious application. Nevertheless, despite the difficulties there are in determining which ones exist and just what they are, some such rules exist in virtue of the status we accord them in our behaviour, loyalties, and deliberations. And clearly it is of the very essence of such rules that they be—within the group whose rules they are—public.

But the rules which we have been constructing in this study are theory-dependent or theoretical. These rules are implied by a given theory. The rules implied by a moral theory are purportedly moral, but they require no publicity or general acceptance. They need not be generally observed, acknowledged, accepted, or referred to. They need only be implied or justified by a given moral theory; and thus they are dependent upon that theory; as theoretical rules they are dependent for their correctness upon the validity of the theory.

Some theoretical and *de facto* moral rules can of course substantially agree; the distinction I am making does not affect that possibility. The theoretical rule against lying implied by a given moral theory, for example, might have exactly the same content as a *de facto* rule against lying. That is, a given theory might agree with current moral beliefs regarding which cases of lying are right and which are wrong. Indeed, this kind of convergence is to some extent the aim of moral theorists. But the two types of rules must none the less be distinguished. That is to say, the grounds

upon which, according to one rule or the other, an act is right or wrong are different.

If we fail to keep this distinction in mind confusions arise. For example, philosophers sometimes ambiguously speak of 'justifying' or 'accounting for' moral rules. The rules to which the proponent of a given theory is committed, as a proponent of that theory, are theoretical rules, e.g. primitive rules, ideal rules, and so on. But the rules with which we are all familiar as part of conventional morality are *de facto* rules, e.g. those the Act-Utilitarian recommends as practical aids, as rules of thumb, as cautionary rules. Moreover, the *de facto* moral rules of a given society cannot obviously be matched by the theoretical rules of the kinds of principles we have been examining. For these theoretical rules are, in principle, completely determinate (open-texture of terms aside). In contrast, our conventional moral rules normally include a *ceteris paribus* condition which cannot be analysed away into a determinate qualifier. There is no *de facto* consensus regarding how conflicts among such rules are to be settled. Confusion can be caused by the fact that *ceteris paribus* conditions are sometimes incorporated into theoretical rules. But in the case of rules derived from principles of the types we have examined, if a rule derives from a strong principle then the *ceteris paribus* qualifier is eliminable; if it derives from a weak principle, then how it can be eliminated—if it can be—depends upon the remainder of the moral theory in question. Even if the moral theory is left somewhat indeterminate, the question still arises whether the particular theoretical and normal *de facto* uses of the *ceteris paribus* condition are exactly equivalent. Only if the *ceteris paribus* condition of rules within a given moral theory is expressly designed to be equivalent to conventional *ceteris paribus* conditions is there a possibility of 'justifying' or 'accounting for' *de facto* moral rules in the sense of deriving or justifying equivalents from the particular theory. But if, for example, the *ceteris paribus* condition of *de facto* moral rules connotes 'generally', or 'usually', or 'probably', instead of 'other things being equal' as I have developed it, then there will inevitably be a gap between derived theoretical rules developed from theories such as we have examined and *de facto* ones.

A second source of confusion is the attempt by some theorists

to provide what may be understood as a theoretical explication of certain *de facto* rules. This is essentially the development of general judgements beginning with selected descriptions such as we have already discussed. Thus the Act-Utilitarian may take the *de facto* rule against killing and transform it into 'Killing is wrong, except when the effects of killing would be better than the effects of not killing, all things considered.' And this is of course no longer our *de facto* rule against killing, which is not at all generally understood to have such an Act-Utilitarian explication. It is another question entirely whether this is a better rule, whether it is one which we ought to observe—that is, whether it is the *really* moral rule, in contrast with our merely conventional one.

A third possible source of confusion is that, within some theories which have been regarded as 'rule-utilitarian', the *de facto* rules of a given society are more or less accorded theoretical status. Thus in Toulmin's theory, for example, acts are right or wrong according as they are so regarded under the moral code, beliefs, and conventions of one's community. But Toulmin's use of *de facto* rules should be distinguished from the Act-Utilitarian's, at least from that Act-Utilitarian position I am now presenting. Whereas the Act-Utilitarian regards them as rules of thumb that one would be well-advised to consult, on Toulmin's account we are logically obliged to consult them—they are logically indispensable.

In response to those practical considerations showing that there is a risk in attempting to appeal to the Act-Utilitarian principle directly, the Act-Utilitarian may recommend those ready-at-hand, relatively simple, conventional rules of 'common-sense' morality—*de facto* moral rules. These are supposed to be reliable guides to right action: they presumably represent and summarize or generalize upon centuries of mankind's experience regarding the usual utility of socially important types of acts. The best means of ensuring that we do what we must want to do as utilitarians, i.e. maximize utility, is to observe these *de facto* moral rules.

But we must also adopt this escape clause: follow the rules, indeed, but not when you know or are quite certain that breaking one will have better effects on the whole than keeping to it. This

escape clause is essential to the Act-Utilitarian's recommendation, and its inclusion clarifies the status of the rules. For these rules are not determinants of rightness and wrongness; they are rules of thumb, practical aids. The ultimate court of appeal is the Act-Utilitarian principle itself: acts are right or wrong according to their relative simple utilities. This point may be obscured by the doubt some utilitarians feel about whether the escape clause can ever reasonably be supposed to be satisfied; for the weight of the accumulated experience of mankind as embodied in the rules is considerable *vis-à-vis* the chance of error, ignorance, and temptation affecting our direct calculations of utility. In other words, some Act-Utilitarians hold that rarely, if ever, are we practically justified in breaking the rules—that we would be taking too great a risk in doing so. Thus the Act-Utilitarian must admit that we may sometimes be unjustified in doing the right act when this involves breaking rules that we have insufficient reason for breaking. Such a position can be bolstered by drawing a sharp line between questions of rightness and wrongness, on the one hand, and questions of praise or blame and responsibility on the other.

But it must be emphasized that such a practical addendum to the Act-Utilitarian's position need not represent a modification of the Act-Utilitarian principle itself. In recommending general observance of the *de facto* moral rules the Act-Utilitarian is merely recommending a way of deliberating and of effectively solving problems of conduct to best ensure conformity to the Act-Utilitarian principle. And thus his elaborated position need not be viewed as hypocritical or paternalistic. He is not committed to a dichotomy of moralities—one, for the learned, philosophical, conscientious, trustworthy, and another, for the 'vulgar'; nor is he committed to keeping his moral beliefs in any way secret. It is open to him to publish his full position, including his practical reasons for recommending *de facto* morality, and to press hard upon whatever features of it need emphasis.

The Act-Utilitarian's aim in recommending the adoption and observance of *de facto* rules is to provide the most practical approach towards maximizing utility, an approach that could reasonably be expected to have some success, not in another world or at another time, but here and now. The Act-Utilitarian

can claim of the rules he recommends what cannot be claimed for any other set in a given community, not even for a set that an enterprising rule-utilitarian might draw up as having, say, the highest acceptance-utility. We *have* these *de facto* rules already in a way in which we cannot have any other set of rules.

But it should also be observed that the rule-utilitarian does not want merely to recommend rules for adoption in the same way. The rules to which he is committed have a different relation to his principle, to the criterion of rightness and wrongness he accepts, to the notion of maximizing utility, than have the *de facto* rules viewed as rules of thumb. The *de facto* rules are brought in as practical guides which (it is claimed) roughly summarize what the Act-Utilitarian principle itself has to say about rightness and wrongness. But the rules with which the rule-utilitarian is concerned are not practical guides, they are themselves determinants of rightness and wrongness. That is to say, one cannot determine the rightness or wrongness of particular acts (on the rule-utilitarian's theory) without consulting the rules themselves. For as we have seen, on such a theory as Ideal Rule-Utilitarianism, a set of rules is judged by its acceptance-utility; and the rightness or wrongness of acts depends upon what the set of rules with the highest acceptance-utility happens to say about them.

Primitive Rule-Utilitarianism is, as it were, half-way between Act-Utilitarianism and Ideal Rule-Utilitarianism in this respect. For the primitive rules may be regarded as determinants of the rightness and wrongness of acts, but they are not indispensable. Although the primitive rules are theory-dependent, what they imply about particular acts can be gotten directly from the general utilitarian principle from which (PRU) is developed. In the case of Ideal Rule-Utilitarianism, however, appeal to the basic governing principle, (IRU), yields only directions for the general acceptance test, and no criterion for directly assessing particular acts.

Now let us consider how the rule-utilitarian may be regarded as exploiting the practical obstacles to maximizing utility faced by (AU). The argument I shall reconstruct cannot be ascribed to any particular moral philosopher, although various versions

of it are suggested in more than one paper on the subject and in current discussions.

The argument against Act-Utilitarianism is sometimes put in this way: (1) The point of Act-Utilitarianism is to maximize utility. (2) But if everyone applied and acted upon (AU), utility would none the less not be maximized. That is, because of the practical obstacles discussed above, some bad consequences would result from everyone's applying and acting upon (AU)—or at least some good consequences that could be produced would not be produced. (3) Therefore Act-Utilitarianism is unacceptable as a moral principle, since its point would be defeated (frustrated) by its general practice. The last point is sometimes made by saying that (AU) is self-defeating or self-frustrating.

Let us call this the *general coherence argument*. I have purposely formulated it in somewhat ambiguous terms (as it often is found) with the intention of contrasting two interpretations. One is mistaken, the other is plausible. The import and limitations of the latter can best be grasped by a brief examination of the former.

But in order to develop this contrast I must first crystallize a distinction implicit in the foregoing discussion, which will become essential as we proceed. The distinction is between *accepting* and what I shall call *following* a rule or principle. Acceptance, as I have said, suggests that one tries to apply and act upon a given rule or principle correctly. But one follows a rule or principle only when one succeeds in correctly applying and acting upon it, i.e. by doing what the rule or principle requires.

Now consider the general coherence argument in its stronger form, in terms of 'following': (1) The point of Act-Utilitarianism is to maximize utility. (2) But if everyone followed (AU), utility would not be maximized. (3) Therefore (AU) is unacceptable as a moral principle, since its point would be defeated by its being generally followed.

This is not, however, an appropriate rendering of the general coherence argument. For the argument is intended to make an issue of the difficulties preventing successful acceptance of (AU), while in this strong version successful acceptance, i.e. following,

is assumed. (2) is necessarily false. But this version is the one sometimes suggested in the literature. Harrod (pp. 147–8) and more clearly Harrison (pp. 131–4) seem to hold that the argument in terms of 'following' would work against (AU) but not against its general utilitarian analogue (GU). For they suppose that the Act-Utilitarian is prevented from taking threshold factors into account, that the Act-Utilitarian cannot acknowledge effects that depend for their production upon general practices, since he is committed to viewing and assessing acts separately, one at a time. But this is of course that central error, that misapprehension about simple *vis-à-vis* general utilitarianism which I have tried to expose.

The general coherence argument becomes more credible when we displace 'follow' by 'accept': (1) The point of Act-Utilitarianism is to maximize utility. (2) But if everyone accepted, i.e. *tried* to apply and act upon, (AU), utility would not be maximized. (3) Therefore (AU) is unacceptable as a moral principle, since its point would be defeated by its general acceptance. The second premiss in this case is plausible, for reasons already mentioned. It is this mode of argument, based upon what we might call the *relative acceptance-utilities* of moral principles, which seems to open the possibility of fundamental criticism of Act-Utilitarianism and the possibility of a more adequate, alternative utilitarian theory. But there is much in it that requires clarification.

The general coherence argument in its most plausible interpretation can I think be characterized as formal, self-referential, and comparative. It is formal in the sense of not employing a substantive criterion on the basis of which principles are to be judged; self-referential in the sense that each principle to which the general coherence test applies is assessed according to one of its own features, namely, its point; and comparative in the sense that a principle is evaluated in comparison with alternative principles.

Consider the notion of formality first. Clearly this is not a consistency test; it is not in that sense formal. Indeed, no one—so far as I know—has claimed that Act-Utilitarianism is self-contradictory. Nor would it seem to be inconsistent in the sense of implying a limited range of incompatible judgements, for its strength is combined with a comparative feature.

It is not uncommon for philosophers to propose necessary, presumably non-substantive tests for moral principles, rules, or maxims. Another condition often suggested, for example, is that moral principles must not include essential reference to particular persons, places, or things. I shall not deal generally with the concept of such a formal test, but I hope that the following observations will shed light upon the limitations of the general coherence argument. Apart from the fact that it does not promise to work simultaneously against Act-Utilitarianism and for some non-primitive form of rule-utilitarianism, it does not seem to be a necessary test in the sense that any acceptable moral principle must pass it.

I want to emphasize that the general coherence argument is formal because it is too easy to slip into a methodologically questionable and viciously circular mode of argument, one which has never clearly been distinguished from the general coherence argument in discussions comparing the forms of utilitarianism. Circularity is a danger primarily because the argument is normally employed by utilitarians intramurally, as part of intra-utilitarian disputes, such as when the rule-utilitarian seeks to show the superiority of his kind of principle. In this context, a careless formulation of the argument may suggest that the criterion according to which principles are to be assessed is utilitarian. That is, it may seem that the argument rests upon a suppressed, higher-order criterion to the effect that principles must be compared on the basis of their relative acceptance-utilities.

If we understand the general coherence argument in this way, however, we seem to be committed to holding that the principles assessed cannot possibly be unjustifiable, ultimate moral principles. For if Ideal Rule-Utilitarianism or Act-Utilitarianism or any other principle rests upon some higher-order utilitarian principle concerning the relative acceptance-utilities of principles, then these principles of right conduct such as (AU) and (IRU) become derivative and secondary. And yet it is normally assumed by philosophers that, if any of these principles are valid, they are ultimate and unjustifiable. They may be supported by argument, but they cannot be derived from some other principles nor justified by reference to them.

Moreover, even within the limited context of utilitarianism, the argument tends to beg the question for such a theory as Ideal Rule-Utilitarianism because it is a general coherence test. The condition supposed is the general acceptance of a given principle. But as we have seen, the Act-Utilitarian can offer a similar, but *non*-general coherence argument. For Ideal Rule-Utilitarianism is at least indeterminate with respect to, or at worst excludes, minimizing-conditions. It can thus be regarded as self-defeating for a wide range of cases in which acceptance of the principle is not general. For in such cases, the point of (IRU), which is presumably also that of maximizing utility, will not be achieved, since following (IRU) will not necessarily result in the best consequences.

This suggests, incidentally, that the general coherence test is not a necessary condition of a principle's being acceptable. For the results of the non-general coherence argument may conflict with the results of the general coherence argument. Alternative conclusions which can be drawn are, first, that both test conditions are necessary (then perhaps no such principle is acceptable), or, second, that the same point cannot be attributed to both principles.

We can avoid viewing the argument as substantive by viewing it as self-referential. A principle can be assessed according to some criterion supplied by reference to the principle itself. The general coherence argument then rests upon the notion that an acceptable moral principle is coherent in the sense that its general acceptance will not defeat its point.

What do we mean by 'the point of' a principle? Baier, for example, uses the notion of a 'purpose': he says that a rule (or maxim) is self-frustrating and therefore unacceptable (unfit to be universally taught) if its 'purpose is frustrated as soon as everybody acts on' it (pp. 196–7). But what are we to count as, and how do we determine, the purpose of a principle? Strictly speaking, a principle (or rule or maxim) has no purpose. It is adopted for a purpose, perhaps, used for some purpose, and so on. It seems odd, however, to attribute a purpose (point, aim, goal, or end) to a principle (or rule or maxim) itself.

This is not a verbal point. Consider two suggestions made by Harrison in his version of the general coherence argument. He

speaks, apparently indifferently, of 'the purpose of the people who applied' a given principle and of 'the ends which determine them to adopt it' (p. 132). But neither of these will clearly do, nor are they obviously equivalents. The purpose of one who applies a given principle, i.e. in applying it, need not be the same as the ends which determine him to adopt it. Moreover, we can have many different purposes, ends, aims, or goals in applying a given principle, in adopting it, and so on. Since we are dealing with the acceptance of purportedly moral principles, however, we might be properly concerned only with high-minded purposes, &c. If so, we can restrict our attention to such as the following. One's purpose might be simply to do right and not wrong, and to try to do this by acting on the basis of a principle which one regarded as moral. (Are there any more appropriate purposes?) One's reason for accepting it may be, not that it is a utilitarian principle as such, or that its point is maximizing utility (one of which Harrison seems to suggest), but that it seems the most viable alternative among the candidates.

But perhaps this does not help in the present case. Perhaps we should begin by assuming that maximizing utility is the point of Act-Utilitarianism and proceed from there. I think then that what must be meant by 'the point of' a principle is the end, aim, purpose, or goal to which one is logically committed in accepting the principle in virtue of its having certain features. When one accepts Act-Utilitarianism as an ultimate, overriding moral principle—which is the only way, it seems, one can fully accept it—one is logically committed to maximizing utility. One is committed to applying the criterion of the best consequences in assessing acts. One is committed to trying always to produce the best possible state of affairs.

But in this sense, maximizing utility can be regarded as the point of (AU) as a consequence of that principle's being a teleological (utilitarian) principle, and, in particular, one of a certain species, e.g. positive (concerning the maximization of utility), and not negative (concerning merely the minimization of suffering, &c.). And there seems no similar respect in which a non-teleological (non-utilitarian) principle can be ascribed a point at all. This entails a further limitation upon the general coherence argument. Such a self-referential test is not applicable to all

kinds of candidate moral principles. Thus, passing the test cannot be regarded as a necessary condition of a principle's being a moral principle or of its being acceptable. Non-teleological principles cannot pass this test, but that is not a mark against them. For they can neither pass nor fail it. As we shall see, however, the fact that the test cannot be directly applied to non-teleological principles does not imply that they are irrelevant to it.

In what sense is the general coherence argument comparative? In the first place, some standard is required for determining whether a given principle is self-defeating. Consider Act-Utilitarianism: that principle is considered to be self-defeating because utility would not be maximized by its general acceptance. But what is it, not to maximize utility in this context? We know what it is in particular cases: to effect a worse state of affairs than could be brought about if we had acted otherwise. Hence the notion of maximizing utility has something to do with the notion of an *alternative*.

It might be felt that a principle is self-defeating (in relation to maximizing utility) merely if better states of affairs could possibly be brought about than those which would be brought about by its general acceptance—in other words, if its acceptance-utility is not a maximum. But this may be too strong a condition, for, after all, it is possible that no principle has a maximum acceptance-utility, that is, that no principle is such that, if it were generally accepted, then the best possible states of affairs would be effected. If we do not consider this too strong a condition, we must acknowledge the possibility that every utilitarian principle (or every principle having the point of maximizing utility) is 'self-defeating'.

A related but somewhat different way of interpreting the test—another way of viewing the comparison to be made—is to assess principles on the basis of differences between their respective following- and acceptance-utilities. This would make the test work against Act-Utilitarianism because the defect of (AU) is supposed to be precisely that its acceptance-utility is lower than its following-utility. In the case of (AU), of course, this is to say that its acceptance-utility is not a maximum. In the case of principles that are not extensionally equivalent to (AU), however, their following-utilities will not be as great as the following-utility

of (AU). None the less, their acceptance-utilities may be even lower. Indeed, this may be the case for Ideal Rule-Utilitarianism and other such principles, as I shall argue. Thus, if we view the general coherence argument in this way, it may again work against every utilitarian principle.

There is a further way of understanding the general coherence argument which I would like to propose. Consider the alternative to general acceptance of Act-Utilitarianism. It is the general acceptance of some other principle that can presumably play the same role as a determinant of rightness and wrongness. Thus, we could compare the extent to which general acceptance of a given principle achieves its point with the extent to which general acceptance of other principles achieves the same point. In the case of Act-Utilitarianism, assuming the point to be that of maximizing utility, we would compare the acceptance-utility of (AU) with the acceptance-utility of alternative principles.

Now there are two ways of determining the range of comparisons to be made. We could, on the one hand, compare the acceptance-utilities only of principles having the same point—in this case, only of principles whose point is maximizing utility. Among the principles to be compared with Act-Utilitarianism, therefore, would presumably be Ideal Rule-Utilitarianism. But it should be observed that, although many other principles do not have the same point as (AU), they can be assigned acceptance-utilities. We could then broaden the range of comparison and compare the acceptance-utility of (AU) with that of any alternative moral principle.

If we pursue the argument along the latter lines, allowing a broad, essentially unlimited class of comparisons, there develops the possibility that no principle having the point of maximizing utility will be graded 'acceptable'. For it is possible that the acceptance-utility of some non-utilitarian principle is greater than the acceptance-utility of every utilitarian principle. This would constitute a paradox of utilitarianism. If it were possible to carry out such a test, it might be worthwhile doing so, just to discover whether such a paradox can be confirmed.

But even if we were to compare only principles having the same point, it does not appear that Ideal Rule-Utilitarianism would prove more acceptable than Act-Utilitarianism. This may be

obscured by the formulation of Ideal Rule-Utilitarianism. For the set of ideal rules is defined precisely as having the highest acceptance-utility. It might seem, therefore, that Ideal Rule-Utilitarianism itself has the highest acceptance-utility and that it necessarily would outshine Act-Utilitarianism in such a comparison. This is not however the case. For the acceptance-utility of the principle, (IRU), need not be as high as the acceptance-utility of the set of ideal rules determined by the principle. We know what (IRU) is—but we do not know the details of the set of ideal rules. The considerations determining this set of rules are extremely complex—more so than those determining the primitive rules of (PRU). Thus one could accept and understand (IRU) without having a very accurate notion of the set of ideal rules and therefore one could make serious errors—understandable errors, no doubt, but errors none the less—in trying to apply (IRU). Indeed, it may be harder to determine the set of ideal rules and thereby the implications of (IRU) in many cases than to determine the implications of (AU) or (GU). Therefore, the principle (IRU) itself may have a very low acceptance-utility, perhaps lower than that of Act-Utilitarianism.

All this makes the position of the Ideal Rule-Utilitarian quite ironical. Since (IRU) promises to be extremely difficult to apply correctly, Ideal Rule-Utilitarianism seems to be a most impractical theory, not the sort of theory which overcomes the practical obstacles to successful application of Act-Utilitarianism. And yet the argument for (IRU) begins with and is based upon these practical obstacles.

Two final comments upon the general coherence argument as it now appears. A question could be raised whether Act-Utilitarianism and Ideal Rule-Utilitarianism can be assumed to have the same point. For the notion of maximizing utility plays a significantly different role within the two theories. In the case of Act-Utilitarianism, the notion of maximizing utility (producing the best consequences) is the criterion applied to acts in determining their rightness or wrongness. But not so for Ideal Rule-Utilitarianism. This criterion is applied instead to a set of rules, and, moreover, in a significantly different way. For while following Act-Utilitarianism will produce the best consequences in every case, following Ideal Rule-Utilitarianism will not necessarily

produce the best consequences in every case. In accepting (AU) one is committed to producing the best effects in each action; but in accepting (IRU) one is not so committed. And, since (AU) and (IRU) are not extensionally equivalent, in accepting (IRU) one is committed to regarding as right some acts that do not have as good consequences as every alternative, or to regarding as wrong some acts that do not have worse consequences than some alternative; but in accepting (AU) one is not so committed.

Moreover, I have strong misgivings about the kind of comparison between principles required by a general coherence argument. The approach implicit in the argument is, of course, radically different from the mode of comparison I have employed elsewhere in this study. What I have done up to now—what it seemed quite natural to do—was compare the actual implications of analogous principles, that is, compare what we can correctly infer from them about right and wrong. On the general coherence argument, however, we are required to consider what might be incorrectly inferred from the principles in question, for the practical obstacles to successful acceptance of (AU) include difficulties in determining what the principle actually implies owing to a shortage of information, and so on. But note the kinds of factors we must consider in making such a test and therefore in assessing the 'acceptability' of a candidate moral principle. We are to consider, in effect, the mistakes we make, the errors to which we are prone, our weaknesses, temptations, passions and cravings, our blockheadedness, ignorance, confusion and stupidity—all of which bear upon what we might actually infer and do, as opposed to what we ought to infer and do according to the principles in question. Theories are not normally compared in such a way. No other kind of theory would be assessed on the grounds implicit in the general coherence argument.

The rule-utilitarian is free, of course, to view the general coherence argument in various ways, as I have suggested. It appears, however, that none of these modes of argument will count simultaneously against Act-Utilitarianism and for a theory like Ideal Rule-Utilitarianism. Alternatively, the rule-utilitarian can take a radically different line from the one I have outlined here. He can disavow any intention of developing an alternative to (AU) that is utilitarian in the sense in which (AU) is. He

can say that his theory is simply designed to accommodate the counter-examples to Act-Utilitarianism and that it need not fit any preconceived notions of a utilitarian theory. But in the first place, (IRU) does not seem to satisfy the counter-examples against (AU)—as I shall re-emphasize in the next chapter. And secondly, to take such a stand is to say, not that Ideal Rule-Utilitarianism or some alternative is a better theory of the same kind, but rather that it is a theory of another kind, and that the other is a better kind. But this remains to be seen, especially since the differences between (IRU) and (AU) still do not account for arguments from justice and fairness.

V

LIMITS OF UTILITY

THUS far we have taken the generalization test as if it implied an appeal to utilitarian considerations only. This was done in order to show that strictly utilitarian principles cannot do the job which some have hoped or thought they could. Thus, despite misapprehensions to the contrary, utilitarian generalization provides in no substantive respect an alternative to simple utilitarianism. Primitive rule-utilitarianism accordingly provides no alternative to simple utilitarianism; and non-primitive forms of rule-utilitarianism do not promise to be very satisfactory.

But the generalization test need not be viewed in this restricted way. If the test is to have the special point we often think it has, a point or force which sets it off from, say, Act-Utilitarianism, then we must bring in factors other than utility. I want now to suggest the relevance of non-utilitarian principles or arguments that can be called forth by the generalization test. These arguments mainly concern justice and fairness.

My presentation will be more speculative than in the preceding chapters. I want to suggest the conditions necessary for an appeal to fairness or justice, to give an account of one such argument that cannot be reduced to utilitarian considerations, and to suggest several different auxiliary arguments and principles that may be relevant.

A. *Arguments from Fairness*

Problems of fairness typically arise in social situations where there is co-operation among individuals. Briefly, where there is co-operative effort achieving some beneficial goal, problems of fairness seem to be of three kinds: (i) the relative distribution of benefits and burdens; (ii) the administration of institutional rules (e.g. problems of impartiality and discrimination); (iii) what I shall call problems of co-operativeness. Where a practice is under criticism and subject to modification, or where it is being drafted

and planned, considerations of type (i) are predominant. But in connexion with our aim of making practices and institutions or methods of co-operation maximally efficient, there is a special sort of problem, (iv), which I shall call the determination of fair procedures. I shall deal mainly with (i), (iii), and (iv), as these are related most directly to the utilitarian considerations we have been examining. They are also clearly interrelated.

Those writers who have associated an argument from fairness with the generalization test have paid almost exclusive attention to the first problem, that of distribution. Broad characterized the general case this way:

> You admit that a certain good result can only be obtained by the co-operation of a number of people. Further, this co-operation involves certain sacrifices on the part of all the co-operators. Lastly, the good aimed at is one which, from the nature of the case, must be enjoyed by all the members of a certain class whether this class be identified with the group of producers or not. The enjoyers may not all be producers and the producers may not all be enjoyers. *E.g.*, if any good results come to the victors in a war they will be of such a kind—national prosperity, feeling of national pride, etc.—that they will *ipso facto* be enjoyed by many members of the victorious nation independently of whether they helped to produce them or not. On the other hand, it is quite certain that many of the producers cannot be enjoyers, because they will be dead or injured for life. A feeling of national pride is *e.g.* a very poor compensation for the loss of both eyes and a leg. Now it may be true that just the same good will be produced whether you co-operate or not, but there is no relevant difference between you and those who join which entitles you to the halfpence without the kicks and them to the certainty of the kicks and the possibility of no halfpence. (pp. 387–8.)

Clearly threshold effects are involved in such cases: the co-ordinated action of a number of people is required for certain effects to be produced. Roughly speaking, the problem of distribution arises because of maximizing-conditions: it is possible to produce the good or prevent the evil by means of a general but not universal practice. And since the acts required in such cases

are usually burdensome, it could be argued that better consequences on the whole would result if the practice was indeed not universal. Thus a utilitarian argument would incline towards exceptions based upon maximizing-conditions, whereas an argument from fairness demands that there be no exceptions. For example, to take a trivial case, but one we have already dealt with, a number of people may refrain from crossing a well-kept park lawn because they know that if everyone—indeed, if not everyone, but many—crossed it, the grass would be damaged. The advantages of the good lawn are or can be enjoyed by all; one can derive pleasure from contemplating it (as in Harrison's case), and it might be used for community-wide events and celebrations. Refraining from crossing it is a minimal burden, but none the less we may suppose that those who refrain would get some enjoyment from crossing it, or that it would be more convenient to do so.

But the lawn would not be damaged if only a few crossed it. Thus some might argue that no harm would come from their walking across the grass, since others are generally refraining. A good lawn plus some pleasure derived from crossing it (in the number of physically allowable cases below the threshold) may be presumed to be better than a good lawn only. To avoid the objection that these few crossings will necessarily have disutility (e.g. by setting examples for others and perhaps upsetting the general practice of refraining, or by creating ill will in the community), and thus that the argument from fairness is *really* utilitarian, we can suppose that the individuals who do not conform but cross the lawn do so in the absence of witnesses. But why should others object to their crossing the lawn when most refrain? Surely not because those who complain think merely of the example that is set (for all the refrainers may be firm in their commitments to a good lawn), and surely not because they think that ill will necessarily arises in the community. The reason they complain about the exceptions taken by the few (which cannot be taken by many without harmful effects) is that the few who cross the lawn (or who propose to do so) have no special claim. This is presumably what Broad means when he says that 'there is no relevant difference between you and those who join which entitles you to the halfpence without the kicks'. There is

indeed the difference between the several acts from a strictly utilitarian point of view: some are performed when others are generally refraining. But no particular individual (we are supposing, in this simplified case) has any special claim apart from this fortuitous position. This position is moreover the consequence of others refraining. The enjoyment of the one who does not produce is *parasitic* upon the efforts or restraints of others.

There is no relevant difference among the individuals if they are viewed in a certain way: as potential or actual producers and enjoyers. It should also be observed that the sort of case outlined by Broad is not restricted to acts which are more or less similar *in vacuo*, from a utilitarian point of view. The acts and agents are classed together primarily in virtue of their relation to the existing practice. The relation of the several agents to the practice usually depends in such cases upon an institutional framework: an organization, community, family, economic group, and so on. And in such settings, the several acts that may be required in the course of co-operation need not be similar in any straightforward way: they need merely be associated as integrated elements in a co-operative scheme.

An argument from fairness that is based upon a principle of *just distribution of benefits and burdens* seems, therefore, to require these conditions: (1) there is a general practice or pattern of co-operative behaviour that involves a number of individuals acting in certain ways and actually achieving a desired end; (2) the end is the production of a good or the avoidance (or prevention) of an evil which could not be achieved without such co-operation or such a general practice, e.g. without the general observance of an acknowledged rule, practice, or procedure; (3) the behaviour is typically burdensome, involving some hardship, frustration, sacrifice, inconvenience, pain, or other loss of benefits to the agent; (4) the benefits produced must be shared; that is, at the very least a number of people have the opportunity of sharing in them and ordinarily do; (5) the total benefits distributed must outweigh the burdens required (or else there would be no good utilitarian point to the practice in the first place); (6) it must be possible to produce the good or prevent the evil without the co-operation of everyone who will enjoy the benefits or who will

have the opportunity of doing so (in other words, there must be an occasion in which a non-universal general practice is sufficient to make the benefits available; or else there would be no occasion in which the foregoing conditions were satisfied and some enjoyers who might produce failed to co-operate). Under such circumstances, this argument from fairness, for a just distribution, demands universal co-operation by enjoyers.

Such an argument allows that, for the sake of a just distribution, less value may be produced than if there were not such a balance between producers and enjoyers. In this respect, considerations of fairness or justice are thought to outweigh some considerations of utility. In this respect, therefore, such an argument from fairness is prima facie *non*-utilitarian.

But this is only a rough formulation. Some allowance must be made for special claims, special hardships, special needs—aside from the fact that some enjoyers (such as children) cannot be expected to produce since they are unable to perform the kinds of acts required.

The imposition of burdens seems central to such an argument. If there were no burdens, it might not be necessary to create a co-operative practice (as we sometimes do by legal means) in order to guarantee that the ends are achieved. The burdensomeness of the acts required normally inclines us against co-operation. But legal measures are not necessarily required, nor should it be supposed that all such co-operative arrangements are based upon a formal understanding reached at a given point in time. The requirement is that a co-operative general practice already exists.

Furthermore, the burdens need not necessarily be uniform or evenly distributed. I have suggested merely that the behaviour required be typically burdensome. Rules against smoking in libraries result in mixed burdens: some find it an intolerable imposition and frustration not to be allowed to smoke; others find it a minor annoyance; still others are indifferent or indifferently content; while others are glad to have the force of public sanctions operating to restrict their smoking. While it is true that all alike are similarly restricted by such rules, it must be admitted that the actual burdens are not necessarily the same.

Such an argument from fairness presupposes that the general practice achieving the end must exist. General co-operation must

actually be achieved. For if there were no jointly produced shareable good, there would be no benefits to be distributed. But what if there is no such co-operative practice; what if our aim of establishing it is not yet realized? It would seem that the argument appropriate in such circumstances is a utilitarian one. Those who fail to co-operate under such conditions may do wrong, but their wrong seems to be different from that of non-productive enjoyers. Here no one is an enjoyer: and that is the problem. The argument in such a case is directed at getting minimal (and not universal) co-operation.

Consequently, the argument from fairness based on just distribution preserves minimizing-conditions. It would be pointless to urge someone to 'co-operate' by performing an act for the purpose of achieving some end that can only be achieved by co-operation, when co-operation is not in fact generally forthcoming. This is obviously true in connexion with informal co-operative enterprises, e.g. our case of several individuals who might start their car again by pushing it, when so many refuse that anyone's pushing (among those who are willing) will be fruitless. Within legal systems, cases connected with minimizing-conditions and analogous circumstances are extremely complex. Particular laws may fail to satisfy the conditions necessary for an argument from fairness, e.g. when inhumane or poorly designed, when they are 'dead letters' and observance would be self-defeating, when they are unjust or unfairly administered. But, though an argument for observance of particular laws based upon fairness may fail, we may also have a general obligation to obey the law that is based on fairness and the relative utility and justness of the legal system as a whole. Thus it *may* be wrong to break a bad law in a generally good system.

The argument from fairness suppresses the utilitarian appeal to maximizing-conditions. This yields a result that is indeed stronger than any form of utilitarian generalization. It must be noted, however, that such an argument is not appropriate in every sort of case in which a utilitarian would want to exploit maximizing-conditions. I shall deal later with promising as another kind of non-utilitarian argument that can hold against maximizing-conditions. At present I want to emphasize that this argument from fairness only applies in certain circumstances:

that is, where there is a more or less informal method of co-operation, or some generally acknowledged rule or practice, the observance of which produces some good or prevents some evil, or where there is some existing law that has a similar utilitarian backing. I am not assuming that the actual method, rule, practice, or law is maximally efficient; it must merely produce more in the way of benefits than it imposes in burdens. There may be cases where some good could be produced, or where the good that is produced (as a result of threshold effects) is not, strictly speaking, a result of co-operation. But in such cases there seems no place for an argument from fairness.

It should also be noted that the argument as I have presented it is not restricted to negative utilitarian contexts, not just to burdens which are imposed for the sake of avoiding or preventing some evil. It covers cases in which a positive good is produced, above some assumed norm. Or better, once co-operation is achieved and benefits are produced, the distinction between utilities and disutilities begins to fade; it becomes less useful. For, once a practice is established and the benefits are enjoyed, the loss of any such benefits would reasonably be viewed as a positive loss (as opposed to the elimination of a positive gain).

Given these general remarks, it is possible to sketch out a principle which can be used to strengthen utilitarian generalization. To put it another way, we can formulate an argument which can be adduced in connexion with the generalization test, one which combines considerations of fairness with factors of utility. I shall present the combination of these two features in the form of a 'deduction'. (The reader may, by reference to the following deduction, compare the results of this discussion with the fate of Marcus Singer's 'generalization argument'. In the Appendix I show how a strictly formal argument from fairness, such as Singer employs, adds nothing to an argument from utility.)

U: If the consequences of everyone's doing a certain sort of thing (including merely co-operating by doing one of a number of complementary tasks) would be the production of a good or the prevention of an evil, the benefits of which are more than sufficient to compensate for all the burdens

that are required by everyone's so acting, then it would be wrong for anyone whose action would contribute to the production of that good or to the prevention of that evil to fail so to act.

JD: It would be unfair, and hence wrong, for anyone who benefits from the production of a good or the prevention of an evil, which is accomplished by a number of persons co-operating in doing a certain sort of thing (or in performing complementary tasks), to fail to co-operate.

JP: If the consequences of everyone's doing a certain sort of thing (including merely co-operating by doing one of a number of complementary tasks) would be the production of a good or the prevention of an evil, the benefits of which are more than sufficient to compensate for all the burdens that are required by everyone's so acting, then it would be wrong for anyone who benefits from the production of such a good or from the prevention of such an evil, to fail to co-operate.

The two premisses are, respectively, a utilitarian one, (U), based upon the considerations outlined above (with an eye to threshold effects), and a principle of just distribution of benefits and burdens. The result is a principle of 'just practice'.

Let us consider some of the difficulties in this approach. In the first place, do we want fairness (just distribution) always to weigh more heavily than utility? Is there no point at which the demand for an identity of the two classes of producers and enjoyers can be outweighed by the disadvantages of everyone's making the effort required or maintaining his restraint? Broad suggests that such a point can be reached:

> there may come a point where it is better that some people should refuse to co-operate although this involves an imperfect distribution, than that they should by co-operating produce a much smaller net-balance of goods though perfectly distributed. (p. 389.)

Now what sort of case could Broad possibly have in mind? I can think of one that is already accounted for by the conditions governing the argument from fairness. That is, the burdens

might be so great compared with the benefits produced that a universal practice yields a net loss. An argument from fairness would not apply in such a case. But what is to be done? It is one thing for potential producers to decline membership in a group which is strictly voluntary. They would presumably lay aside any claim to be enjoyers. But membership in some groups (e.g. in a community—citizens) is not quite so voluntary or so easily dismissed. It is also not always possible to cease to be an enjoyer.

Part of the solution to this problem requires that we consider another aspect of the notion of fairness, that of fair procedures. I want to sketch one case to illustrate this facet of the problem. Suppose that a college of one thousand students has a cafeteria that can serve fifteen hundred. The cafeteria provides a congenial atmosphere and good food at the low prices that students can afford. But the costs of operation are such that, if it does not operate at capacity, services must be cut and prices raised. Let us assume that all the students have their meals there, thus accounting regularly for one thousand places. Guests of students are to fill the remaining five hundred places.

Now if every student were to invite one guest, bad consequences would result, for the cafeteria would become terribly overcrowded, unable to serve everyone, unable to serve anyone properly. On the other hand, if no student had a guest the results would also be bad. Suppose that no student has any more reason at the outset than any other to claim a right to invite guests; suppose also that all wish to do so (because of the good food, low prices, and atmosphere). There are no relevant differences among the students before they get in line, say, or before some, with their own guests, fill the fifteen hundred seats.

It would be absurd to say simply that no one should have guests (or that everyone should have guests) because the results of no one's (everyone's) having guests would not be as bad as the results of everyone's (no one's) having guests. To argue either way is to miss the point. It would also be too risky for each student to decide whether to invite guests depending upon how many others do so. For if no one invited guests, or if everyone invited guests, or if everyone calculated for himself, avoidable bad results would occur. What is required instead is an arrangement whereby the facilities could most efficiently be used. What is

required is the establishment of a fair procedure for determining which students at which times may bring guests, a procedure designed to maximize utility without infringing upon the equal claim with which each student begins. A rotation system could be established: this group would be entitled to invite guests at one time, that group at another. The details are not important here. What is important is that the procedure be fair.

Notice the following about such a case. Relevant differences among the students arise out of such a procedure. Under a rotation system, for example, the various students would have determinate claims that could be verified. Secondly, the appeal to utilitarian generalization to determine what any individual should do before the fair procedure is established is inappropriate. The question is not 'Should I (or should I not) invite a guest?' but rather 'How can we set things straight?' The responsibility in this case is, moreover, not primarily that of the individual student, but that of the school administration, student council, or other more or less official body. And after the system has been established, the answer to the individual student's question, 'Should I invite a guest?' presupposes considerations of fairness and not just those of utility.

The significance of fair procedures for our purposes is in the fact that, in contrast with the argument based on just distribution which tends to *suppress* consideration of maximizing-conditions, this approach is aimed at *exploiting* them fairly. Suppose that seven of us have spades and six holes must be dug. The rational approach might be, not to distribute the burdens (which might be difficult to do), but e.g. to draw lots and let the lucky one of us off for the day, thus maximizing utilities. This sort of case should be distinguished from that in which one of the potential producers has an injured back (a disability) and thus has a special claim not to produce.

But this is only a partial answer to the question originally raised, about utility perhaps outweighing fairness. (Here one does not outweigh the other: fairness is applied in order to maximize utility.) I mentioned before that the existing method of co-operation, rule, practice, or law need not be maximally efficient. As was suggested in the case of laws, such co-operative enterprises—any *de facto* institution—may be more or less

humane, just (in its rules), and fair (in its administration). This suggests that a generally useful practice may be subject to criticism of such diverse kinds and to such an extent that one's obligation to co-operate can be outweighed by other moral considerations. But these problems are extremely complex, and I can only suggest a small part of that complexity. Consider a tax law, a favourite case for the proponents of the generalization test. To ask, 'What would happen if everyone failed to pay his taxes?' may obscure the issues. Why are the taxes so high? Are they proportionate to one's ability to pay? Is allowance made for a minimum—indeed a comfortable—standard of living? Is the revenue used so that all share in the fruits, or better, so that those who need most gain most? Are there special loop-holes (such as capital gains provisions) which in effect place a heavier burden on the lower income taxpayer than he realizes, a heavier burden than he would accept if he realized the true nature of the law? How fairly is it administered? And there are also questions like: How can the law be changed? By peaceful and conventional means—or by peaceful and unconventional means—or otherwise? What will the effect of opposition to—or disobedience of—this law be on other, more useful, more efficient, more humane, more just and fairly administered aspects of the legal system? And so on.

It is no use our trying to simplify matters here; simplified arguments will soon prove in practice to be short-sighted and naïve. In any actual case an argument from fairness will most likely occur within a more or less inefficient, more or less inhumane, more or less unjust and unfairly administered set of rules. There is no simple, readily definable set of calculations through which we can run in order to determine whether we should co-operate, instead of doing something else. (The alternative to co-operation is not simply a failure to co-operate; many sorts of actions are open in either case.)

Because of such factors, it might be best to say that one's obligation to co-operate is not perfect or absolute, but rather prima facie. For there may be conditions—not just utilitarian considerations now—which justify a refusal to co-operate in a practice which satisfies the conditions, outlined above, for the application of an argument from fairness.

But even to speak of a prima facie obligation is perhaps to beg an important question. Let us consider the nature of this argument as I have presented it. To begin with, why should such an argument be considered non-utilitarian? It might be claimed that disutilities systematically follow from acts which can be characterized as 'unfair' or 'unco-operative', and thus that the net-balance of good is always reduced when such acts are performed. I have already suggested two kinds of reasons which could be used to support such an attempt to 'reduce' fairness to utility. By setting an example of unco-operativeness, the enjoyer who fails to produce may actually harm the practice through his influence over others ('If he can get away with it, why not me?'); and ill will and dissension, which themselves have disutility, can be created in a community of producers. No doubt other factors could be added. But every such consideration, it seems to me, subtly presupposes the wrongness of the unfair behaviour on independent grounds. None the less, the utilitarian will try to account for our sensitivity to or disapprobation of unfairness as, e.g., a useful psychological and social mechanism serving to reinforce useful behaviour.

There are several reasons for rejecting such a reduction. First, I think it would be hard to show that, in every case involving secrecy—or better, merely lack of publicity, disutility follows. Second, some of our reactions presuppose considerations of fairness. For example, one who violated the rules of a fair procedure and thus failed to co-operate would be subject to criticism for acting wrongly, not because his act was harmful (for it need not be), but because it was unfair.

In general, it is necessary to distinguish between our reasons for criticizing certain modes of behaviour, on the one hand, and the consequences of that behaviour, on the other. We need not suppose that unco-operative (or unfair) behaviour is not harmful. It can be disastrous in its long-range effects. But reasons based upon utility are only one relevant kind of reason.

The claim that unco-operativeness necessarily has disutility should not be confused with a superficially similar claim which more effectively reduces the argument based on just distribution to utilitarian considerations. The first argument was designed to show that unco-operative behaviour has disutility of the ordinary

kind, that is, that it has instrumental disutility leading to a loss of intrinsic good. The second argument is based on the contention that just distribution is itself an intrinsic good and that unjust distribution is an intrinsic evil. In other words, on this approach 'just' and 'unjust' are the primary words, to be applied to states of affairs, i.e. to distributions of goods or to the relation of the class of producers to the class of enjoyers; whereas these words (or 'fair' and 'unfair', 'co-operative' and 'unco-operative') are only derivatively applied to acts in so far as they are conducive to one type of distribution or the other. This accommodation of fairness to utility has been called 'ideal' utilitarianism, a term which is intended to signify that values other than pleasure are admitted. (This usage should not be confused with that of 'Ideal Rule-Utilitarianism'; the 'ideal' in the latter has nothing to do with value-criteria, while it has in the former.)

Broad's approach, like that of the other writers who have dealt with fairness in relation to the generalization test, is 'ideal' utilitarian. He asserts that:

> the goodness or badness of a complex state of affairs is not a function merely of the goodness or badness of its parts. A certain set of goods distributed in one way between a number of people may constitute an intrinsically better state of affairs than the same set distributed differently. And the appeal to 'fairness' seems to rest on the principle that the best possible state of affairs is reached when the group of producers and that of enjoyers is as nearly identical as possible. In fact common-sense would probably go further than this and say that the best possible result was reached when (*a*) producers and enjoyers are identical and (*b*) the share in the good produced that falls to each producer is proportional to his sacrifice in producing them. (p. 388.)

I think this accounts for part of our notion of justice or fairness. Whether we accept this as a reduction of fairness to utility will depend to some extent on terminological matters. For example, it seems to make perfectly good sense to speak of fairness (or justice) outweighing utility—which could not happen if fairness (or justice) were simply an aspect of utility. But this obstacle to the reduction could perhaps be overcome by noting that fairness

(or justice) is a special kind of intrinsic good, a property only of very complex states of affairs that have certain features.

It should also be observed that Broad's notion of just distribution—benefits strictly commensurate with burdens—is only one among several plausible notions. Broad's criterion may be contrasted with this alternative, for example:

> From each according to his resources,
> to each according to his need.

(See Colin Strang, 'What If Everyone Did That?' *Durham University Journal*, xxiii (N.S., Dec. 1960), p. 8.) Broad's criterion would run, analogously:

> From each according to his benefits,
> to each according to his burdens.

No doubt neither constitutes by itself a sufficient criterion. But these suggest the shortsightedness of that form of egalitarianism according to which all persons are to benefit (or share) equally. It would seem that some consideration must be given to burdens and needs.

There is another difficulty: how to assign credit, as it were, to individual acts for being productive of a just or unjust state of affairs. I take it that any effects which are produced must ultimately be assignable to one or more individual acts. If just (or unjust) distribution is a kind of utility, then particular acts must be spoken of as producing, or contributing to, such a distribution. And presumably acts will be wrong in so far as they contribute to unjust distributions (having that kind of bad effect), or right in so far as they contribute to just distributions (having good effects). But this implies a number of difficulties. If the principle of just distribution of benefits and burdens requires an identity of the classes of producers and enjoyers, the producers perhaps enjoying a share commensurate with their burdens, how do we view the following cases? (i) a producer who is not an enjoyer (perhaps a dead soldier on the winning side); (ii) a producer who does not enjoy a share commensurate with his burdens (perhaps a maimed soldier on the winning side); (iii) a would-be producer who gives his all for the losing side. These are rather gruesome examples, but they serve to illustrate the following point. On the 'ideal' utilitarian view, where only

a balance between producing and enjoying is considered, the producers in the first two cases must be regarded as producing intrinsic evil in so far as they produce without enjoying. To that extent *their* acts are to be regarded as wrong. And in the last case, no credit at all is to be given to the would-be producer who is unable to produce through no fault of his own. The first two cases illustrate the fact that, even on the 'ideal' utilitarian account, there can be a conflict of fairness (or justice) with (ordinary) utility: for the acts of such producers have both wrong-making and right-making characteristics which are their unjust distribution-efficacies and their utilitarian-efficacies. All three cases indicate, moreover, that this approach represents an inversion of our reflective attitudes in such matters. For we do not ordinarily think that the producer who fails to enjoy is doing anything wrong. We may regard him rather as the victim of an unjust system or as a victim of unfortunate and perhaps disastrous circumstances. And we regard the would-be producer who makes the effort, in some circumstances at least, as doing the right thing or something worthy of our admiration.

These consequences of 'ideal' utilitarianism can be avoided if we allow ourselves to consider the motives or reasons for which individuals perform the acts they do as grounds of their rightness and wrongness. That is, we must consider reasons and motives independently of the utilities or disutilities connected with them. We must also allow ourselves to consider fully the identity of the enjoyers, to ask, in other words, in whose interests acts are performed. It seems clear that in one use of the term 'fair', an unfair act is primarily one performed by an individual who tries to get (for himself) something for nothing, who tries to avoid contributing while he consumes, who tries to take advantage of the efforts and restraints, sacrifices and burdens, hardships and inconveniences of others. The question relevant to fairness is, then, not whether one's act produces good or bad effects, but whether one acts in the kind of way I have just generally described.

This may be put in another way. The utilitarian criterion is usually assumed to be applicable, not to all human acts or actions, but only to that proper sub-class of acts or actions which are sometimes called 'voluntary' or 'intentional'. (It is not part of my argument that these two terms mean the same.) For we sometimes

do things unintentionally or involuntarily (not 'of our own free will'), and such acts are supposed not to be properly subject to assessment as 'right' or 'wrong'. (See for example, Moore's discussion in *Ethics*, e.g. chaps. i and vi.) Thus we may understand the utilitarian's general concern for intentions, reasons, and motives very roughly in the following way. He requires that an act be intentional, and thus, in effect, that it be performed (from the agent's point of view) under *some* more or less determinate description. But he does not assess acts on the basis of their intended effects, only on the basis of their actual effects. (Or perhaps their probable effects, but here 'probable' should not be confused with 'intended'.) Thus, the utilitarian is not especially concerned with *the* description under which the agent performs his action, i.e. with *how* the agent views his act, nor with *why* the agent performs it (except as motives or reasons happen to have utility or disutility in some way), nor with what the agent expects or hopes to 'get out of' doing what he does (if he expects or hopes for anything). For the utilitarian, such considerations are not relevant to rightness and wrongness.

Yet in the argument from fairness that cannot be reduced to 'ideal' utilitarian considerations, we are concerned precisely with the agent's motives and reasons and intentions in a special way. It makes a difference whether I view my act as merely crossing the lawn, or as crossing it on the supposition that others are generally doing so, or as crossing it when I know that others are not generally crossing and that I could not get the pleasure I get from crossing it if they were not generally refraining. We are concerned with, among other things, whether *the* description of the act under which I perform it coincides (more or less) with the complete relevant description. Unfair behaviour in this connexion is, roughly, performing the act in circumstances satisfying a maximizing-condition, with the intention of doing so for the benefit of oneself, when one knows (or has good reason to believe) that one's personal benefit is only made possible by the efforts and restraints, sacrifices and burdens, hardships and inconveniences of others.

If such intentions, motives, or reasons for acting are relevant to the assessment of acts as right or wrong independently of the

consequences produced, then utilitarianism is an inadequate doctrine, in any form.

It is tempting to try to give a utilitarian account of this aspect of fairness on the basis of conditions of human life. Human life as we know it—not in any particular form, but in general—could not exist or be maintained without a good measure of co-operation and co-operativeness. There could not be a community of parasites, or even of would-be parasites, because such behaviour presupposes social life in which others are co-operating. Thus it might be thought that the utility of fairness is that it is essential to social life itself. But in any particular case this will not suffice for an argument against unfair behaviour—especially not against trying to take advantage—because universal co-operation or co-operativeness is not essential to the maintenance of social life or of specific social institutions. A good deal of parasitic and would-be parasitic behaviour can be absorbed without appreciable damage to the usefulness of such institutions. In other words, on a utilitarian account of fairness allowance must be given to threshold considerations, to factors such as how many others are being co-operative. And in some cases the utilitarian would be obliged to argue for unfairness (or unco-operativeness) for the sake of maximizing the over-all net-balance of good. But the significant feature of the present argument for fairness (based on motive)—as opposed to the 'ideal' utilitarian or the more conventional utilitarian argument—is that it calls for universal co-operativeness. No one can be justified in exploiting, or in trying to exploit, others.

Thus, it does not seem possible to account for this aspect of fairness by assimilating it to utility. Fairness in this respect seems directly related to the notion of social, as opposed to anti-social, behaviour. One who acts unfairly is to that extent not a responsible human being, not a responsible member of society, not a social creature.

B. *The Practice Conception of Promising*

The topic of promising has been accorded a central place in discussions of the nature of morality in general and of utilitarianism in particular. I propose, in conclusion, to discuss some

relevant aspects of the problem in connexion with moral rules and the limits of utilitarianism.

To begin, let us return to our examination of the Act-Utilitarian's characteristic attitude towards moral rules. As I have already indicated, the Act-Utilitarian views moral rules as rules of thumb, practical aids that are useful in deliberation. One exponent of this approach, J. J. C. Smart, has put it in the following way:

> In practice the extreme [i.e. Act-] utilitarian will mostly guide his conduct by appealing to the rules ('do not lie', 'do not break promises', etc.) of common sense morality. This is not because there is anything sacrosanct in the rules themselves but because he can argue that probably he will most often act in an extreme utilitarian way *if he does not think as a utilitarian*. ('Extreme and Restricted Utilitarianism', *Philosophical Quarterly*, vi (Oct. 1956), p. 346.)

The last clause (which I have italicized) may be misleading. I suggest that it be understood as, 'if he does not deliberate, in each and every case, as he might be encouraged to do by his acceptance of extreme utilitarianism, i.e. by calculating simple utilities directly'. Smart argues that one is obliged in practice to rely heavily upon the *de facto* (conventional) rules of morality because it is either unwise or impossible to do otherwise, owing to the complexity of circumstances and effects, our usual lack of information, the pressures of time, our temptations arising out of personal interest, and so on (pp. 346–8).

Smart's argument is not intended primarily as an analysis of moral rules or of social rules in general, but is intended rather to expose the 'superstitious rule-worship' which he attributes to rule-utilitarians (p. 349). His targets are those conventionalistic rule-utilitarians who, on the basis of a quasi-logical argument, have elevated *de facto* moral rules into a special place in ethical theory; who hold that, in so far as *de facto* moral rules cover particular cases, the rightness and wrongness of acts are strictly determined by reference to such rules, even when the acts required by them would not have the best consequences, even when we know that by acting otherwise we could produce more good. Such writers hold that it is inappropriate (a logical mistake)

to appeal to utilitarian considerations directly, and that assessments of particular acts must be mediated whenever possible by appeal to general rules that have some grounding in utility.

Smart argues, on the contrary, that the rightness of individual acts is to be judged only

> by their consequences, and general rules, like 'keep promises', are mere rules of thumb which we use only to avoid the necessity of estimating the probable consequences of our actions at every step. The rightness or wrongness of keeping a promise on a particular occasion depends only on the goodness or badness of the consequences of keeping or of breaking the promise on that particular occasion. (p. 344.)

I want to suggest, and criticize, two attitudes towards or theories about *de facto* social rules which may be read into the foregoing passage. One of these is simply implausible; it amounts to an inadequate view of rules in general. The second is mistaken, but in a different way—not as an analysis of *de facto* social rules, but rather as a defective normative doctrine regarding the sufficient conditions for justifiably breaking *de facto* rules. This distinction is important because the defects of the first approach that might be attributed to the Act-Utilitarian are exposed merely by a logical analysis; whereas in the second case a substantive or normative analysis is required.

Let us consider the first interpretation. The Act-Utilitarian may be saying that all social rules are merely rules of thumb, practical aids useful in determining the likely utility of acts. Smart suggests this extension of his argument in the following way. First, he demurs at separating moral from non-moral *de facto* rules (see pp. 353–4). Secondly, Smart treats a particular non-moral social rule (a rule which is not part of the moral code of any community) in the same way he treats *de facto* moral rules. He wants to show that, just as in the case of the *de facto* rules against lying and breaking promises, the non-moral 'rule of the road' provides no 'logical *reason* for action but an anthropological *datum* for planning action' (p. 353). He argues that there are cases in which it would be irrational, because so likely to be disastrous, to follow the 'rule of the road' (keeping, say, to the right) and in which one would therefore be justified in breaking it.

Let us view this argument as an attempt to explicate, on theoretical grounds, the real import of such rules. Rules like 'do not lie' (or 'lying is wrong'), 'do not break promises' (or 'breaking promises is wrong'), or 'keep to the right', may be viewed as elliptical formulations of *ceteris paribus* rules. The Act-Utilitarian's argument can be understood as claiming that the suppressed *ceteris paribus* condition can in each and every case be taken in terms of the (*AU*) *escape clause*: 'except when you know or are quite certain that the effects of not following the main body of the rule would be better'. The *ceteris paribus* condition can be viewed in this way on the basis of Act-Utilitarianism and some precautionary advice about its application.

It is prima facie plausible to view *de facto* moral rules in this way because the grounds for claiming justifiable exceptions to moral rules are themselves moral considerations only. That is to say, once one accepts a given moral theory as completely general and adequate in every case, one may view the exceptions which are supposedly allowed by the elliptical moral rules as determinable on the basis of that moral theory. The considerations which justify using the *de facto* moral rules to begin with, and those which determine justifiable exceptions to them, are all of one piece. If one accepts Act-Utilitarianism one can view the *de facto* moral rules as if they have the (AU) escape clause as an implicit qualification.

But one cannot do the same sort of thing with many other social rules. This applies, for example, to the rule of the road, tax laws, rationing restrictions—such rules as are part of a complex legal system within which legally justifiable exceptions can be determined. For example, we cannot reasonably understand and think and act as if the legal rule requiring that one pay *n* per cent. of his income in taxes, *really* goes: 'Pay *n* per cent. of your salary to the government in taxes, except if you know or are quite certain that the effects of not doing so would be better.' Exceptions which are allowed to this legal rule are in a strict sense determined, not by such an (AU) escape clause, but by the legal system of tax laws, appeals, administrative action, and court decisions. How much is in fact owed to the government is determined by such means. If exceptions are allowed to a given rule, they are allowed by the system of which the rule is a part.

If utilitarian considerations play any role in determining such exceptions, they do so because they are legally incorporated into the tax system as part of some other qualifying tax laws, in the course of administrative action, or in the realm of court decision in which there is some room for judicial discretion.

It is a logical point (with some social consequences) that not all rules may be regarded as rules of thumb and that we cannot reasonably regard most social rules as subject to our personal moral-theoretical qualifications. But the Act-Utilitarian need not be understood as making the mistake we have attributed to him on the first interpretation of Smart's view. He need not contend that the (AU) escape clause indicates the exceptions actually allowed by the rule or system of rules in question. He might, alternatively, view the clause as determining the exceptions one is justified in making to a given *de facto* rule. The (AU) escape clause may be viewed, in other words, as indicating the purportedly sufficient moral conditions for breaking any social rule. Such a view is fully consonant with Act-Utilitarianism, for the clause summarizes the grounds that, to an Act-Utilitarian, seem sufficient for breaking a given rule.

It is a mistake to think that the (AU) escape clause correctly summarizes the sufficient conditions for breaking *de facto* rules, but this mistake is different from the elementary error we have just examined. This mistake stems from utilitarianism in general: the Act-Utilitarian ignores obligations to observe rules when those obligations are not grounded in utility. The position of the Act-Utilitarian is inconsistent with the arguments from fairness, for example. When we act within the context of a useful, co-operative, *de facto* practice, we are simply not free to act as if the rules constituting that practice did not exist, as if we played no role within the practice, as if our enjoyment of the fruits did not entail special duties and obligations (as well as rights) that cannot be reduced to utilitarian considerations. One is not justified in breaking a rule just because one knows or is quite certain that the effects would be better if one did so. The latter may be a necessary condition for justifiably breaking a *de facto* rule. It may even be the case that, whenever one is justified in breaking such a rule, in refusing to co-operate, better effects will

flow from one's refusal than from one's co-operating. But the Act-Utilitarian escape clause is not the only factor to be considered.

My argument thus far draws heavily upon John Rawls's distinction between the 'summary' and 'practice' conceptions of rules. ('Two Concepts of Rules', *Philosophical Review*, lxiv (Jan. 1955), 3–32.) But our positions are different. Rawls's argument is, I think, significantly incomplete; a gap in it resembles the one I have indicated in the Act-Utilitarian's argument. Rawls's argument is moreover misleading with respect to promising. I shall expand upon these points briefly.

The first conception of social rules attributed to Smart is manifestly what Rawls refers to as the summary view, whereby

> rules are pictured as summaries of past decisions arrived at by the *direct* application of the utilitarian principle to particular cases. Rules are regarded as reports that cases of a certain sort have been found on *other* grounds to be properly decided in a certain way (although, of course, they do not *say* this). (p. 19.)

But not all rules may reasonably be regarded as rules of thumb. Some rules, for example, serve to define a practice. Actions falling under the rules of a practice can only be understood (or can only be fully understood) in the context of the practice, and in the terms of the rules themselves. The very descriptions of such acts make essential reference to these rules. One could not, for example, speak of promising or of keeping or breaking a promise outside the context of rules and conventions in which certain forms of behaviour count as (are taken as, are mutually understood as) the making of promises. One can swing a stick at a ball, but one cannot strike out except when playing baseball—which amounts to acting under a system of rules. Similarly, one cannot owe or pay taxes outside a system of taxation. Indeed, one cannot even give away money or pay for goods outside a system of conventional exchange, currency, property, and so on.

Let us now examine Rawls's application of this distinction. His immediate objective is to show a way between two horns of a dilemma in moral philosophy. The Act-Utilitarian insists that the (AU) escape clause can always be applied to determine when one is justified in, say, breaking a promise (breaking the rule

against breaking promises). And the critics of utilitarianism have contended that this is simply not allowed by the practice of promising and thus that utilitarianism is so far inadequate. Rawls points out that the several parties in their arguments and counter-arguments 'fail to make the distinction between the justification of a practice and the justification of a particular action falling under it' (p. 16). He seems to argue, moreover, that because certain kinds of acts (the descriptions of which make essential reference to the rules of a practice) can be performed only within the context of such a practice, there is no legitimate appeal to considerations not specified by the rules of the practice. The rules of the practice of promising simply do not allow an unlimited direct appeal to utility. Certain exceptions based on utility are admittedly allowed—one can be justified in not keeping a promise in order to avoid extremely bad consequences—but these are allowed *by* the rules of the practice of promising.

But on the other hand, the practice of promising itself can be defended on utilitarian grounds at a different level: not by consideration of each and every case of promising and keeping or breaking promises, but by consideration of the usefulness of the practice as a whole.

> Indeed, the point of the practice is to abdicate one's title to act in accordance with utilitarian and prudential considerations in order that the future may be tied down and plans coordinated in advance. There are obvious utilitarian advantages in having a practice which denies to the promisor, as a defense, any general appeal to the utilitarian principle in accordance with which the practice itself may be justified. (p. 16.)

Thus two sorts of questions must be distinguished: (1) internal questions, about acts covered by the practice, e.g. 'How can I do *A* (make a promise, make a will)?' 'What act is required under these circumstances (once the promise is made)?' The answers to these are determined by appeal to the rules of the practice; (2) external questions, about the practice as a whole, e.g. 'How useful is the practice, given the rules as they are?' Considerations of utility as such are inappropriate in determining answers to questions of the first kind, but they are in order in the second case. Thus, the fact that the (AU) escape clause cannot be

appended to rules in the practice does not entail that the ground of one's obligation to keep one's promises is not utilitarian. In so far as the basic justification of the practice of promising is utilitarian, the programme of utilitarianism is saved.

Let us suppose for the present that there are rules of the practice of promising which indicate when one is justified in breaking a promise; that there are rules to which we can appeal in determining whether a particular promise ought to be kept, all things considered. It does not matter how we think of such rules, but we can suppose that there is a general rule of the form, 'It is wrong to break one's promises, except . . .' Let us suppose further, as Rawls along with the critics of Act-Utilitarianism supposed (and I think we must agree with him), that there are some cases in which promises ought to be kept, all things considered, although the effects would be better if they were not. We may even suppose that some promises ought to be kept (that it would be wrong to break them) when the promisor knows or is quite certain that the effects of breaking the promise would be better than the effects of keeping it. On Rawls's account, it would seem that (some) such promises ought to be kept because the rules of the practice require that they be kept, and because the practice is generally useful.

Consider now those cases in which keeping a promise according to the rules governing promising would have worse effects than breaking the promise (and thereby breaking the rules). An Act-Utilitarian may be pictured as arguing, along the lines I have attributed to Smart, that breaking the promise in such a case is *really* justified, (1) on the first version of Smart's view, because the exception is really allowed by the rules (assuming the (AU) escape clause to be part of the rules), or (2) on the second version, because the circumstances amount to conditions which are sufficient to justify breaking any rule. On Rawls's account, these reasons are misguided and inadequate. The Act-Utilitarian's argument fails because he misunderstands the import or nature of the rules in question. The Act-Utilitarian offers utilitarian reasons for deciding what to do in a case covered by the rules of the practice of promising. But such utilitarian considerations are relevant only to external criticism of the practice as a whole;

they are irrelevant to the question whether the particular promise should be kept.

Even if we allow this much (the core of Rawls's argument), he is not able similarly to rebut an alternative utilitarian argument. An external, rule-utilitarian criticism of the rules of the *de facto* practice of promising can be correlated with every such abortive simple utilitarian criticism. The abortive criticism shows that in some cases better results would come from breaking the rules than from keeping them, all things considered. But this entails that for every such case there is a class of exceptions that might be allowed by the rules. If such exceptions were allowed by the rules the results would be better on the whole. That is to say, better results would come from observing an alternative set of rules that allowed exceptions in such cases; the best results would come from observing a set of, roughly, primitive rules. (They would most likely have no minimizing-conditions, for reasons already suggested.) The *de facto* set of rules governing promising is defective in so far as it excludes any exceptions that would be allowed by the alternative primitive rules. It is defective from a strictly utilitarian point of view: better results would flow from the observance (in general, or in particular cases) of a different set of rules. But if this is so, and if external assessments are based upon utility, then the criticism of the rules cannot be dismissed on logical grounds. Since the practice as a whole is subject to utilitarian criticism, what reason is there for observing the rules—at least in those cases in which observance of the rules conflicts with the directions of utility? It is fully consistent with Rawls's position for a utilitarian to argue that one is justified in breaking the *de facto* rules just in those cases in which they diverge from the alternative primitive rules. For in just these cases the *de facto* rules have no utilitarian backing. (The alternative is not to observe the rules at all.)

It should not be supposed that the rule-utilitarian's rejoinder can be answered by referring to the disutility of the alternative practice. The alternative practice is, *ex hypothesi*, most useful in general and in every particular case. It cannot be said, for example, that general confidence in the sanctity of promises would seriously be impaired under the alternative system, or that we could not as effectively plan. For such an alternative system

must be presumed public: everyone would know (as they presumably know now) the kinds of conditions which justify the breaking of promises and so could plan accordingly. Perhaps more promises would be broken under the alternative system, but it is not as if no promises could justifiably be broken under the present system. And the justified broken promises would have good effects.

The fact that the alternative system does not exist is also irrelevant, under the conditions of the argument suggested by Rawls. For the question is not whether we should try to institute the alternative system. The question is, rather, whether the *de facto* system can fully be supported on utilitarian grounds. Since it is vulnerable to utilitarian criticisms, our reasons for observing the rules as they are seem weakened. We cannot suppose the present system above criticism. Promising is merely a *de facto* practice; it can have defects; therefore it can be criticized. Moreover, it cannot be claimed that the rules of *de facto* practices can never justifiably be broken. They can be broken. One is sometimes justified in breaking the law, for example. The question is: when, under what conditions, is breaking the rules of a given *de facto* practice justifiable? This is the issue that the rule-utilitarian criticism forces us to acknowledge. This is the issue that might be obscured by Rawls's arguments based merely upon the distinctions between summary and practice rules and internal and external questions.

There are two sides to this new issue. First, what reasons have we for observing the rules of a *de facto* practice? In the present case this becomes, what reasons have we other than the usefulness of the practice—for if the practice is not maximally useful, it is subject to utilitarian criticism. Secondly, what conditions must be satisfied for justifiably breaking the rules of a *de facto* practice? In the present case this becomes, what conditions other than criticism on utilitarian grounds must be satisfied, for if only this condition had to be satisfied, we would be justified in breaking rules much more often than we normally think we are.

My general contention is, in other words, that there is a gap in the argument implied by Rawls. Utilitarian considerations are not the only relevant ones in such cases; they will not suffice. It matters not whether we apply the test of utility directly, to

individual acts falling under the practice, or indirectly, to the general utilities of acts falling under the practice, or again indirectly, to the rules of the practice. As our arguments for extensional equivalence should have suggested, it will not help in rebutting a simple utilitarian objection merely to offer a rule-utilitarian argument—unless the latter is specially constructed.

It would be tempting now, in view also of our discussion of fairness, to attempt to use that notion to heal the breach. My arguments thus far may suggest, for example, that the reason we have for keeping promises (when the results would be better if they were broken) is based upon an argument from fairness. The promisor benefits from a useful, co-operative, *de facto* practice and must accordingly contribute by keeping the rules when his turn comes. Such an argument would work in other cases. But I shall now argue that it will not work here.

Consider first what the utilitarian's attitude towards promising must be. Presumably he can and does make promises as things stand, in the present system. He can understand that in making a promise he incurs an 'obligation' in this sense: he has altered his relationship with others, he has encouraged expectations in others which did not exist before, and bad effects would come from his failure to keep his promise. The promisee's expectations and plans and personal arrangements would be frustrated and upset, there would be some loss of confidence in the practice of promising, the promisor's reputation might be damaged, and so on. Thus, he can understand his act of promising as establishing an 'obligation' in the typical utilitarian sense that he now has good reasons for keeping and against breaking his promise, reasons based solely on the relative utility of acting one way or the other.

But is this enough? What does the utilitarian think he is doing when he makes a promise? When he says 'I promise' (or does something else to the same effect) does he think he is saying (or does he mean): 'I promise, but'—(i) 'before I keep this promise I shall review the situation and then determine whether or not the effects would be better if I break it; and if so I'll break it' or (ii) 'if this case falls into a class of cases for which justifiable non-fulfilment would be allowed under an alternative system, &c.,

then I will not keep it'? Perhaps we should dispense with the reference to promising itself, and thus his promising would signify: 'Yes, I'll *do* it, but'—(i) 'before I *do* it . . .' or (ii) 'if this case . . .'. Now if the utilitarian means something like that, he had better say so, for the promisee (unless he is a very similar-minded utilitarian who has been briefed) will not take it that way. The promisee will understand—will have been given to understand by the promisor's promising—that the promisor is committed to some future performance; that, if something unexpected and unforeseen happens which makes it impossible or very harmful for the promisor to keep his promise, the promise need not be kept, and so on. There are no broad utilitarian escape clauses understood. The commitment is by no means tempered by a view to the assessment of the practice.

The utilitarian could make such conditions explicit, and (if they were accepted) that would constitute a kind of promise. We might say that it is not a firm commitment. On the other hand, he might make the promise without letting on that there are hidden strings attached. If he does so, and if it is merely the greater good in which he is interested, then we would say that his behaviour is grossly misleading and irresponsible, perhaps that he is a hypocrite. There is a fine line between the universalistic utilitarian's escape clause and the egoistic utilitarian's: in the latter case, if the promisor exploits the promising situation to his own advantage, by means of a bit of artful deception, he is subject to very severe criticism. And it would be wrong to suppose that such criticisms against either kind of utilitarian in such a case would concern only blameworthiness and not at all the wrongness of the acts.

The point is, surely, that this criticism applies equally to the Act- or rule-utilitarian. To suppose that utilitarian considerations of either (or any) kind could unrestrictedly govern the rules of a practice of promising is simply to confess a lack of understanding of what the making of a promise is. I am not of course denying that promising is useful—that would be absurd. But it is not simply because keeping promises in particular cases or in relevantly similar cases is always (or generally) useful that one has an obligation to keep one's promises. It may be that humans would not have ways of making promises and thus would not

have promising if promising were not a useful form of activity—though it seems strange to suppose that humans might not have a way of making promises or, more generically, of coming to understandings involving such commitments about future performances. But this is not to say that, given the useful human institution or form of activity called promising, our only (good, moral) reasons for keeping our promises are, directly or indirectly, utilitarian.

There are many gaps in the utilitarian's account of morality. These typically result from his failure to take seriously rights, duties, and obligations which are not exclusively grounded in producing good or preventing evil. The gaps in some of his arguments—those especially associated with the generalization test—can often be filled by appeals to justice and fairness. But we should not try to force fairness and justice to do all the jobs that must be done to complete the account of morality. We should not make the same sort of mistake the utilitarian makes, that of assimilating many different kinds of reasons and considerations in morality to one kind.

Now reconsider Rawls's account of promising. His argument suggests that there are gaps in the strict utilitarian account, but he does not make that clear, and he does not indicate what the gaps are. First consider his general account of practices. He suggests that internal and external questions are all the relevant questions there are. But there is an overlapping question: what to do when one is involved in a practice which itself is subject to serious criticism on utilitarian or other grounds. Rawls implies not only that utilitarian considerations may suffice for external criticism of a practice, but also that one's reason for keeping the rules of the practice is simply that the practice has (some) utilitarian backing. But no *de facto* practice is likely to stand up under exhaustive utilitarian criticism. If utilitarian grounds are our only reasons for supporting the practice, those become considerably shakier when the practice is imperfect.

But if special rights, duties, and obligations arise as a result of one's participation in a useful, co-operative practice, where one is enjoying its benefits, then the gap can be filled. One could claim, for example, that it would be unfair to fail to assume one's burden when one's turn came. In fact Rawls suggests

something along these lines in his general remarks about practices:

> Practices are set up for various reasons, but one of them is that in many areas of conduct each person's deciding what to do on utilitarian grounds case by case leads to confusion, and that the attempt to coordinate behavior by trying to foresee how others will act is bound to fail. As an alternative one realizes that what is required is the establishment of a practice, the specification of a new form of activity; and from this one sees that a practice necessarily involves the abdication of full liberty to act on utilitarian and prudential grounds. (p. 24.)

Consider how and why one's liberty is abdicated. One has an obligation to co-operate in a useful, co-operative, *de facto* practice. But certain conditions must be satisfied for this obligation to have effect. One's liberty is not abdicated unless there is a shared good produced, unless it is fairly shared, unless special claims are met, and so on. The fairness obligation is relative to the conditions and circumstances of the practice itself, and not merely to the individual's acts within it. It may be regarded moreover as an external obligation to keep the rules. These rules require certain determinate modes of behaviour.

This account is not fully applicable to promising. For what is important and distinctive in the case of promising is that one is obligated to do only what one promises to do, and just because one promises to do it. The relevant obligation in the case of promising is not something grounded in the conditions of the rules of the practice and of enjoyment and production under them; it is grounded in the act of promising itself. One abdicates one's liberty, not in respect of some set of rules as a whole, but rather in respect of each and every promise when and only when it is made.

In drawing a contrast between the 'practice' of promising and other practices, it is worth noting how on one account promising has been distinguished from useful, co-operative, *de facto* practices. In his discussion of the various special rights which are created by 'voluntary actions', H. L. A. Hart separates these cases. ('Are There Any Natural Rights?', *Philosophical Review*, lxiv (April 1955), pp. 183–6.) Rights are created or conferred on

promisees 'by the deliberate choice of the party on whom the obligation falls', i.e. the promisor, in virtue of a particular 'voluntary transaction between the parties' (pp. 184–5). Other special rights arise, in contrast, out of a 'mutuality of restrictions':

> when a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions when required have a right to a similar submission from those who have benefited by their submission. (p. 185.)

This is very similar to the case of useful, co-operative, *de facto* practices, save that Hart does not explicitly impose a condition of utility on the joint enterprise. I take it that Hart's treatment of this second case of a special right (with correlative obligation) does not require a 'deliberate choice of the party on whom the obligation falls'. It would seem, rather, that the sort of 'voluntary action' involved here is acquiescence and willing enjoyment. In the case of promising, considerably more is normally required.

But the difference between the two sorts of cases is not simply a difference of degree in the subtlety of 'agreement' normally presupposed. The difference is, among other things, that one sort of right and obligation (fairness) presupposes a full-blown practice with various kinds of rules, including rules requiring certain modes of behaviour, whereas the other sort of right and obligation (promising) does not. Turn once more to the logical aspect of Rawls's analysis. He indicates that certain kinds of acts simply cannot be done outside the context of a practice. It is logically impossible for such acts to be performed, for their very descriptions presuppose certain rules. In the case of promising, only one type of rule (or convention) is presupposed. We might call this an *enabling* rule, in virtue of which certain patterns of behaviour constitute the making of a promise. We are enabled to make promises because such conventions exist. If we make promises, we cannot help but act in accordance with the enabling rules.

But certain features of this situation should warn us of dangers ahead. Unlike the making of wills (which is determined by legal enabling rules), the making of promises admits an indefinite number of ways in which such an understanding can be reached. The enabling rules are very informal. They are, in this respect,

not unlike the rules that allow for an indefinite variety of grammatically correct combinations and permutations in a language such as ours. There never will come a day when nothing new can be said—or at least said differently. Informal rules are typically open-ended and allow expansions and new discriminations at the borderlines. In this respect they are unlike the rules of law (which allow only a minimal open texture) and the rules of arithmetic. We should not expect the rules of the 'practice' of promising to be as easy to formulate as those of law and arithmetic.

Now consider a second kind of rule within practices like punishment (Rawls's other example), taxation, military service, and rationing. There are a number of species which we may for our present purposes group together and call *requiring* rules. These tell us, for example, to pay so much in taxes, to register for selective service, to do such and such if we hold this or that office, and so on. They may not 'tell' us in so many words, 'Do *A*.' Legal rules often indicate legal wrongs or offences by indicating penalties for certain kinds of behaviour—for failing to do, in other words, what the legal rules require.

First it should be noted that Rawls's logical point does not have the same force in connexion with requiring rules as it had with enabling rules. Whereas it is logically impossible to fail to observe enabling rules when doing some acts, it is perfectly possible to fail to observe requiring rules. Failing to pay one's taxes and failing to register for selective service (as required) are quite possible. Moreover, they are not necessarily wrong acts just because their descriptions presuppose some reference to the (enabling) rules of the practice they contravene. Whether they are not just legal wrongs, but are wrong, all things considered, is a question which cannot be answered simply by reference to law.

More important, however, is this: for all practical purposes, there are no requiring rules in the 'practice' of promising. We cannot appeal to rules that indicate the necessary and sufficient (or just sufficient) conditions for justifiably breaking promises; we cannot appeal to rules that indicate exception-making criteria or that list exempting-conditions; we cannot appeal to an arbitrator (judge) or administrator or lawyer for advice or a ruling. We cannot decide whether one is justified in not keeping his promise in such a way, we do not do so, and we need not.

The closest we get to a requiring rule for promising is a vague, implicitly *ceteris paribus* 'keep your promises', or 'do not break promises', or 'promises ought to be kept', or 'it is wrong to break one's promises'. . . . Various individuals also have somewhat diverse general expectations and feelings regarding which circumstances can be taken as constituting conditions of justifiable non-fulfilment. Rawls acknowledges this to some extent:

> It must, of course, be granted that the rules defining promising are not codified, and that one's conception of what they are necessarily depends on one's moral training. Therefore it is likely that there is considerable variation in the way people understand the practice, and room for argument as to how it is best set up. For example, differences as to how strictly various defenses are to be taken, or just what defenses are available, are likely to arise amongst persons with different backgrounds. But irrespective of these variations it belongs to the concept of the practice of promising that the general [i.e. unlimited simple] utilitarian defense is not available to the promisor. (pp. 30–31.)

It is a significant part of what I am arguing that the last claim would better be made with respect to the concept of *a promise*, as opposed to the concept of *the practice of promising*, and further that in speaking of the concept of a promise we are not necessarily making implicit reference to certain requiring rules of such a 'practice'.

Perhaps Rawls's comments can be understood as claiming that the requiring rules governing promising are as extraordinarily flexible as the enabling rules. Thus we are merely unable to 'pin them down'. But consider what we do when we try to determine whether a promise ought to be kept (all things considered), or whether one is justified in not having kept his promise. We do not appeal to rules: we argue, debate, deliberate, criticize, explain, justify, and so on. In actually determining whether a given case of promise-breaking is justifiable or not, we are usually deciding a kind of case which has never been decided before, in which our previous determinations are useful only up to a point, as analogical aids in determining what it would be reasonable to conclude in the given case. In actually determining

what is justifiable and what is not by means of deliberation, explanation, argument, and so on, we are as much creating, expanding, qualifying, modifying, and determining the 'rules' governing promising as we are appealing to them.

In sum, then, I have two criticisms of the approach suggested by Rawls. First, it implies that utilitarian considerations suffice to account for our obligations to observe the rules of useful, co-operative, *de facto* practices. We find, however, that such obligations seem to be effective even when the practice is not maximally efficient, even when it may be subject to utilitarian criticism. And we also find that a certain conception of fairness can bridge the gap.

I can suggest one possible source of this error. In discussing the utility of a practice, Rawls seems to make a mistake which is also often made by utilitarians who seek to 'justify' or 'account for' our obligations to observe *de facto* rules. The mistake is, on the surface, merely to note that the rules are useful, and to leave it at that. But this involves applying an unspecified 'principle of utility'. (Note that Rawls suggests that the same principle of utility could be applied to acts as well as to practices; but it is not at all clear how such a principle would be formulated.) Thus we are not told *how* useful a rule (or practice) must be if it is to be 'justified'—somewhat useful, more useful than not, rather useful, maximally useful . . .? (And we are not told whether the utility accrues to the rule or practice as it is normally, i.e. imperfectly, observed, or whether the test concerns the usefulness of the practice or rule were it very generally followed.)

Now it should be clear that the sort of utilitarian principle required here by the utilitarian must be strong and comparative. This suggests that justification of a rule or practice on utilitarian grounds requires maximum utility—that it be the most beneficial practice possible, or (on negative utilitarian grounds) that it be most effective in minimizing suffering and inconvenience. But as we have observed, *de facto* practices and rules are rarely, if ever, maximally efficient in this sense. (While, on the other hand, an argument from fairness does not presuppose maximal efficiency.) It would seem that, if the principle of utility to be applied had been specified, it might have been realized that a whole-hearted utilitarian blessing could not

be given such practices or rules on either Act- or rule-utilitarian grounds.

Secondly, I am arguing that in any event the concept of a practice as a system of requiring as well as enabling rules is not applicable to promising. This criticism should be kept distinct from the first.

It must in fairness be noted that I have attributed a mode of argument to Rawls—as I have said, it is suggested or implied by what he actually says and fails to say—which might not accurately reflect his position. In a later paper ('Justice as Fairness', *Philosophical Review*, lxvii (April 1958), 164–94), Rawls indicates that the argument which he applied to promising (the one we have been examining) was primarily intended to show that

> there is a peculiar force to the distinction between justifying particular actions and justifying the system of rules themselves. Even then I claimed only that restricting the utilitarian principle to practices as defined strengthened it. I did not argue for the position that this amendment alone is sufficient for a complete defense of utilitarianism as a general theory of morals. ('Justice as Fairness', p. 168, fn. 5.)

Moreover, in that later paper (pp. 179–83) Rawls develops a conception of a duty of fair play which is closely related to what I have dealt with, must less adequately, under the headings of 'fair procedures' and 'co-operation (or co-operativeness)'. Rawls indicates the relation of this conception to the special rights and obligations which Hart claimed arose in conditions of 'mutuality of restrictions'. Most important for our purposes, Rawls also explicitly makes the kind of distinction I am urging when, for example, he says:

> The duty of fair play stands beside other prima facie duties such as fidelity and gratitude as a basic moral notion; yet it is not to be confused with them. These duties are all clearly distinct, as would be obvious from their definitions. ('Justice as Fairness', p. 181.)

Furthermore, one of the major, explicit points of Rawls's argument in the later paper is that utilitarianism cannot fully account

for this duty, that is, that there is a fundamental limitation upon utilitarianism in general.

Thus, it is not clear that we can attribute the strictly utilitarian argument which is suggested in the 'Two Concepts of Rules' paper to Rawls. But this is not the sole, nor perhaps the main, question at issue. I have argued, not only that there is a gap in the utilitarian account, but also that the game/practice model which is developed by Rawls and applied by him to promising cannot profitably be so applied.

And there is a pressing reason for emphasizing this point. The merits of Rawls's distinction between the two concepts of rules and his application of the game/practice analogy have been generally acknowledged. But the limitations which he sought to place upon that method of analysis have not fully enough been appreciated. Thus, in the first place, Rawls found it necessary to reiterate in the later paper (as cited above) that he disclaimed any intention of applying the model to morality in every respect and in general, and thus of reconstructing and salvaging utilitarianism. (See 'Two Concepts of Rules', p. 32, fn. 27.) Nevertheless, it is not uncommon to find Rawls characterized as a 'rule-utilitarian' and therefore praised or blamed accordingly. But more important than this general misapprehension about Rawls's intentions are others' applications of his analysis—applications which are suggested by Rawls's own treatment of promising, or the duty of fidelity. For the game/practice model has been applied by philosophers to all aspects of morality, including morality as a whole. But the special obligations—for example, those of fidelity, gratitude, and fair play—cannot themselves be viewed as analysable in terms of game-like practices. If practices are significantly relevant at all—as, in the case of fair play, they surely are—they are not relevant as the rules in accordance with which we, say, criticize others' behaviour as being unfair or unfaithful or ungrateful; they are relevant, rather, as the settings or contexts within which such obligations arise. Thus our duty of fair play arises out of our participation in useful, co-operative practices—but there is not a further practice, the rules of which determine when we are or are not to call a course of action 'unfair'. And yet the rules of the useful, co-operative practices themselves are not concerned with fairness. The rules

of *de facto* practices require that we perform, or refrain from, acts of this or that kind, under these or those circumstances. Considerations of fairness, gratitude, and fidelity are, in contrast, critical considerations, not subordinate to or determined by any *de facto* rules (not even the rules of conventional morality). Thus, when we apply such notions as fairness, gratitude, and fidelity in morals, we are not performing moves in a game. We are rather concerned and occupied with the basic critical and self-critical operations of morality itself—of morality, as opposed to conventions and codes.

APPENDIX

IN *Generalization in Ethics* Marcus Singer attempts to deduce a form of utilitarian generalization from a simple utilitarian principle. Considerable interest has been aroused by Singer's argument. Critical comments have accumulated at an extraordinary rate, mostly centring upon flaws or ambiguities and possible equivocations in the proposed deduction. My intention is least of all to survey the commentaries and present a critique along similar lines. I propose instead to take this opportunity to apply the analytic machinery developed in the foregoing in order to examine facets of Singer's argument that do not otherwise promise to be illuminated within the limitations of strictly critical studies.

THE PATHOLOGY OF AN ARGUMENT

1. *The principle to be deduced, the generalization argument, GA*

> If the consequences of everyone's doing x would be undesirable, then it would be wrong for anyone to do x.

This is not Singer's most frequent formulation of GA, though it seems to be one of a number of allegedly equivalent formulations. (See fns., pp. 65–66.) For example, the consequent is otherwise rendered as, e.g., 'not anyone (no one) has the right to do x' and 'no one ought to do x'. Although the choice of terms can be crucial in the course of the deduction itself, I think it is clear that the formulation I have chosen is fully compatible with Singer's intentions, and it makes GA patently a negative non-comparative form of utilitarian generalization.

But Singer does not just change the formulation of the consequent. He also elaborates, modifies, and patches up GA to enable it to do all sorts of jobs and to meet apparent counter-examples. (See, e.g., pp. 68, 72–73.) Thus, on the basis of Singer's explicit instructions and actual usage, a full-blown formulation of GA would seem to be something like the following:

> If the consequences of every member of K's acting or being treated in a certain way would be undesirable on the whole, while the consequences of no member of K's acting or being treated in that way

would not be undesirable on the whole, then no member of K ought to act or be treated in that way (it would be wrong for any member of K to act or be treated in that way) without a reason or justification.

I am going to ignore some of these elaborations, for the following reasons: (1) We can ignore the '. . . being treated . . .' aspect of GA because it adds unnecessary complications, it is not incorporated in the principle which is to be deduced, and Singer does not emphasize it. (2) We can ignore the 'every member of K' stipulation since that is covered by conditions of relevance and our inclusion of descriptions of the agent in descriptions of the 'action'. (3) We can ignore the 'while . . . whole' qualification, which incorporates Singer's hedge against 'invertible' applications of GA, because that is added out of an unwarranted fear that incompatible judgements could be justified by reference to GA. In general, the possibility of generating inconsistencies is ruled out by the weakness of GA; in particular, a whole range of unnecessary and only apparent anomalies can be prevented merely by applying GA with respect to complete descriptions only.

But we cannot similarly ignore the final qualification, 'without a reason or justification'. This I shall take to be a straightforward *ceteris paribus* condition, the equivalent of 'other things being equal', an interpretation which is borne out by Singer's application of it. We cannot ignore the condition, although it is suppressed in his deduction and in the formulation which we have given GA above, because it plays a central role in the deduction itself, and an understanding of its genesis will be essential to an understanding of the deduction.

2. *Preliminary view of the deduction*

The principle of consequences (C) states that: If the consequences of A's doing x would be undesirable, then A does not have the right to do x. The following principle (GC) is what I called a generalization from C: If the consequences of everyone's doing x would be undesirable, then not everyone has the right to do x. Now the generalization principle (GP) may be stated as follows: If not everyone has the right to do x, then not anyone (no one) has the right to do x. The generalization argument (if the consequences of everyone's doing x would be undesirable, then no one has the right to do x) clearly follows from GP and GC. (p. 66.)

There are several obvious difficulties in this argument. First, what is the status of the premisses? Second, C is ambiguous. Although it is clearly a simple utilitarian principle, Singer allows for two significantly

different interpretations. Third, GC is not in any apparent sense truly a generalization from C. What is its link with C? Fourth, GC is itself ambiguous and perhaps the main source of trouble in the deduction. Finally, the whole argument is elliptical, is backed up by misleadingly inadequate comments, and is, moreover, pointless. I shall deal with each of these points below.

But before examining the deduction by means of its several parts, let me sketch the essentials of the argument as a whole. Clearly, we must try to render this deduction as at least a valid argument, unless in so doing we force upon Singer some obvious absurdity (as opposed to an understandable mistake, misapprehension, slip, or blunder).

There is no doubt that Singer viewed the total argument in two stages. First (pp. 34–46, 64–65) he argues that the premisses are 'necessary'—thus to prove that the conclusion not only follows from the premisses but is also true (or, in Singer's terms, 'necessary'). As part of this first stage GC is somehow derived from C. Second, as Singer clearly indicates, GA is deduced directly from GP and GC: it is supposed to 'follow from' these and only these two premisses. Therefore, the deduction itself (the second stage of the establishment of GA) is apparently supposed to have the form of a hypothetical syllogism:

(I)

GC: If p then q
GP: If q then r

GA: If p then r

This is the form of argument which Singer employs. But it is not strictly the form he requires. For (I) is elliptical, since the *ceteris paribus* clauses of GP and GA are suppressed. The full deduction would be:

(II)

GC: If p then q
GP: If q and s then r

GA: If p and s then r

This is also a valid form of argument. Here 's' represents the *ceteris paribus* condition which is placed, as I have recommended, in the antecedents of the respective principles it qualifies. Placing it in this way allows us most easily to reassemble the full deduction as a valid argument.

Now if the actual deduction is itself to be valid there must be no equivocation. These conditions must be satisfied: (1) the antecedent

of GC, i.e. '*p*', must be identical with the antecedent of GA; (2) the consequent of GC, i.e. '*q*', must be identical with the antecedent of GP; (3) the consequent of GP, i.e. '*r*', must be identical with the consequent of GA; and (4) the *ceteris paribus* condition, i.e. '*s*', must not be a source of equivocation.

We need not, I think, be troubled about the *ceteris paribus* condition. It has as it were no substance in GP; for the conditions of relevance are determined by the substantive principles being applied. Since only negative non-comparative utilitarian principles are here involved, there is no basis for supposing any equivocation upon the *ceteris paribus* condition. Thus we can ignore this clause to some extent in what follows, and treat the deduction as if it were in fact (I), as Singer actually formulates it, instead of (II), for which (I) is elliptical. But we cannot totally ignore it, as we shall see.

Our approach, then, will be to render as far as reasonably possible GC and GP so that the three conditions (1)–(3) can be satisfied.

3. *The formal premiss, the generalization principle, GP*

> If it would be wrong for someone to do *x*, then it would be wrong for anyone to do *x*.

Again, this is not Singer's favourite formulation. But if we are to derive GA from GC in terms of 'wrong', then this sort of formulation is required.

GP is the pivot of the deduction. This is quite clearly in line with Singer's view, for as he indicates, the 'transition' from 'some' to 'all' which is mediated by GP is 'essential to the generalization argument' (p. 5). We can understand this as also implying, in view of the course of the deduction, that the 'transition' is essential to the deduction itself. GP provides the link between GC, the consequent of which is given as 'not everyone has the right to do *x*', and GA, the consequent of which is given as 'no one has the right to do *x*'. It is plausible to view this as a transition from 'some' to 'all' since 'no one has the right to do *x*' is general and 'not everyone has the right to do *x*' may be regarded as particular. For 'not everyone has the right to do *x*' may be understood as 'it is not the case that everyone has the right to do *x*'; and this in turn becomes 'someone does not have the right to do *x*'—as I shall argue.

Viewed in the way I prefer, in line with my earlier formulations of utilitarian generalization, the consequent of GC and the antecedent of GP become 'it would be wrong for someone to do *x*' (particular), while the consequents of GP and GA become 'it would be wrong for

anyone to do *x*' (general). As I shall later show in more formal terms, this can easily be viewed as a transition from 'some' to 'all'.

The sense of GP is this. The place-holder '*x*' must stand for a kind of action. (Thus '*x*' is a place-holder for descriptions in GC and GA as well.) According to Singer, if it would be wrong for someone to do *x*, it is because there is a reason that can be adduced against doing that sort of thing. Doing *x* would violate a moral rule or principle. But the reason (rule, principle) applies equally to all instances of *x*. Hence, if it would be wrong for someone to do that sort of thing (i.e. if it would be wrong to do that sort of thing in at least one case, which necessarily involves some particular person doing it), then it would be wrong for anyone to do so (i.e. it would be wrong for anyone to do that sort of thing in any case).

Let us examine GP more closely. It might seem that the move from GC to GA is a significant one; that is, that it yields a principle (GA) which implies more than, or at least something other than, what C or GC implies. For it is a very common supposition that similar simple and general utilitarian principles have different implications. Singer's remarks about the 'transition' from 'some' to 'all' suggest something of this sort. It is as if from a (negative) simple utilitarian principle one could infer that a particular act is wrong, because of its undesirable consequences, while GP enables us to 'generalize' by putting the act into a class all members of which must then be counted as wrong (which GA seems to imply). We might put this another way, as Singer clearly indicates: so far as we are concerned with a simple utilitarian principle like C, any acts to be counted as wrong must have undesirable consequences (that is a necessary condition); but so far as we are concerned with a general utilitarian principle like GA, acts can be counted as wrong even though they do not have undesirable consequences (it is no longer a necessary condition). For Singer says that

> the generalization argument does not imply that the consequences of each and every act of the kind mentioned would be undesirable. By reason of the generalization principle it implies that each and every act of that kind may be presumed to be wrong. Yet from the fact that an act is wrong it does not follow that its consequences would be undesirable. (p. 67.)

The last assertion is no doubt true. If it stood alone it would not be problematic. But Singer seems to imply that an act which is counted as wrong *by GA* need not have undesirable consequences. This is either false or true and elliptically misleading. It depends on how we take 'wrong': as 'wrong *sans phrase*' or 'presumed to be wrong' (i.e. 'prima facie wrong'). If Singer means the former, then he is mistaken.

For GA in its strong form is merely the general utilitarian analogue of C and thus is extensionally equivalent to the latter; in its weak form it is not extensionally equivalent to C, but it implies no more—it implies exactly what C implies, but in a weakened, *ceteris paribus* form. If Singer means 'presumed to be wrong', as the passage suggests, then his last assertion is misleading. GA can admittedly be used to generate good reasons against (presumptions against) acts which do not themselves have undesirable consequences. Such reasons or presumptions would be based upon significantly incomplete descriptions in respect of which the acts have undesirable tendencies. But such reasons or presumptions necessarily can be rebutted because acts which do not have undesirable consequences on the whole cannot have undesirable tendencies, all things considered. Thus some of the acts which we can presume to be wrong on the basis of GA—those acts which we cannot call wrong on the basis of C—cannot ultimately be counted as wrong by GA.

But since it apparently seemed to Singer (as it has to others) that analogous principles were not extensionally equivalent, the 'transition' from 'some' to 'all' which is mediated by GP seemed to be significant. This imbues GP with a good deal of importance in the deduction. But what actually is GP? Singer himself identifies it as what 'has traditionally been known as the principle of fairness or justice or impartiality' (p. 5). Clearly GP is that formal principle which I have called the principle of generality, a principle which tells us something about moral reasoning but nothing whatsoever about its substantive grounds.

To speak of *a* or *the* principle of generality may be misleading. What we are appealing to here may perhaps best be viewed as a consequence of the maxim, 'Treat like cases alike.' GP may be regarded as a particular instance or implication or codification of one aspect of that maxim. An indefinite number of such principles can be formulated in different contexts: some about right and wrong acts, or about acts that ought or ought not to be done, or about rights that people do or do not have, and also about praise and blame, about virtues and vices, about desirable and undesirable states of affairs. In other words, in every different context of moral reasoning there are general considerations; in all such contexts like cases should be treated alike.

Indeed, Singer formulates what he calls 'the generalization principle' in several different ways, not all of which are equivalent. GP formulated in terms of 'wrong' (or perhaps in terms of 'ought to do' or 'has the right to do') is merely a formulation suited to the job at hand. So long as its purely formal character is preserved, there are no limits to how we can formulate such a principle.

Let us now consider the full formulation of GP with the *ceteris paribus* condition made explicit:

> If it would be wrong for someone to do x, then it would be wrong for anyone to do x without a reason or justification.

In our idiom this is equivalent to:

> If it would be wrong for someone to do x and other things are equal, then it would be wrong for anyone to do x.

The qualification is important, for if GP must be so qualified, so must GA. The *ceteris paribus* condition is necessarily transmitted from GP to GA in a valid deduction.

Why must GP be so qualified? To understand Singer's reasons for adding the *ceteris paribus* condition we must understand something of the workings of his moral system. GA is to play the major role as the ground of moral rules and reasons. But GA is to be applied according to the method of rebuttals: not necessarily with respect to complete, but generally with respect to partial, descriptions. Thus, the *ceteris paribus* clause weakens GA and thereby weakens the judgements derivable. This allows such judgements to be rebutted; the rebuttal is a further 'reason' or 'justification'. In effect, the function of the *ceteris paribus* condition in Singer's system is to allow for the possibility that there may be relevant differences among acts which are somewhat similar but not necessarily similar in all relevant respects.

Now as I have already indicated, two possible systematic uses of the *ceteris paribus* condition must be distinguished: (1) to prevent anomalous conflicts between judgements derived from different principles within a single system, so as to preserve consistency within it; (2) to prevent anomalous conflicts between judgements derived from a single principle when that is applied according to a method of rebuttals. The first is reasonable: it is necessary in a heterogeneous moral theory. But the second is totally unnecessary and can be misleading. We can place a premium upon completeness of descriptions and, recognizing that in practice we may overlook significant features of an act, treat the inferences which we make from a given principle as subject to correction. The judgements we infer are implied by the given principle only if the descriptions are relevant and complete. But our actual inferences can be mistaken. A mistaken inference, a judgement that we derive from a principle on the basis of a significantly incomplete description, can be viewed as a mis-play, it can be discarded. We need no recourse to *ceteris paribus* conditions.

There is none the less something gained by weakening GA. For GA is non-comparative, and therefore if it were strong it would yield

incompatible judgements in 'lesser of two evils' cases. But such a fundamental defect in a principle—an internal inconsistency—is no good reason for weakening it. What is required is a comparative principle instead.

Let me relate this to GP. It seems that Singer thinks mainly in terms of partial (perhaps significantly incomplete) descriptions as substitution-instances for the '*x*' in GP as well as in GA. But there is no need to treat GP in this way. By doing so, Singer implies that the main body of GP represents only 'treat like cases alike' in its narrowest meaning; as if the *ceteris paribus* condition adds the qualifier, 'but of course treat relevantly different cases differently'. It seems reasonable to suppose, however, that one cannot understand (the main body of) GP or (the main body of) the general maxim from which it is derived unless one also understands that relevantly different cases are to be treated differently. When the maxim (or such a principle as GP) is actually applied, the criteria of relevance in accordance with which it is applied must include provision for relevantly distinguishing as well as relevantly associating. If we assume, on the other hand, that relevantly complete descriptions are the substitution-instances of '*x*' in GP (except in mis-plays), we have no need for a *ceteris paribus* qualification. If we retain and transmit it to GA, we not only obscure the anomalous features of GA but also encourage the illusion that GA condemns acts which do not have undesirable consequences—the illusion that the 'transition' from 'some' to 'all' which is mediated by GP really achieves something we do not already have.

4. *The substantive premiss, a 'generalization from' C, GC*

> If the consequences of everyone's doing *x* would be undesirable, then it would be wrong for someone to do *x*.

This, again, is not Singer's favourite rendering; but it is the one best fitted for the deduction as we are reconstructing it. In the present case, however, the choice of terms is critical, and so I shall explain my choice.

Singer also uses these alternative consequents in GC which he seems to regard as equivalents:

> not everyone ought to do *x*
> not everyone has the right to do *x*.

In rendering GC so as to yield a valid argument, I am in effect transforming these into:

> someone ought not to do *x*
> someone does not have the right to do *x*.

Not only are these three consequents not equivalent to begin with, but the transformations do not preserve their original meanings.

We have a choice here. However we formulate GC, we shall have to make similar adjustments in GP and GA to preserve the validity of the deduction. Some of these reformulations will yield implausible versions of GP and GA. I shall not deal with all the possibilities, but only briefly with those which seem consonant with Singer's intentions and which preserve validity. In reconstructing the deduction in terms of 'wrong', I can assimilate GA to the other forms of utilitarian generalization. In reconstructing GC in these terms I can indicate the importance of GC and locate the central error in Singer's reasoning, an error related to Singer's failure to deal with threshold considerations. I also want to claim this, that 'not everyone . . . not' and '. . . someone . . .' can be understood as 'at least one person (in at least one case) . . .'. If we proceed in this way, the deduction becomes:

GC: If the consequences of everyone's doing x would be undesirable, then it would be wrong for someone to do x.

GP: If it would be wrong for someone to do x [and other things are equal], then it would be wrong for anyone to do x.

GA: If the consequences of everyone's doing x would be undesirable [and other things are equal], then it would be wrong for anyone to do x.

This is a valid argument.

The significance of GC lies in the fact that it is apparently intended to take account of threshold effects. The hint that this is what Singer was vaguely suggesting is given in the following passage:

> Thus GC has as its consequent 'not everyone ought to do x', instead of 'everyone ought not to do x', because supposedly if not everyone does x the undesirable consequences that would result from everyone's doing it would be avoided. (pp. 66–67.)

This implies a reference to threshold effects because those are the effects that depend for their production upon the general practice of some acts. Singer's remark can be understood in the following light. The general practice of x is a necessary and sufficient condition of the undesirable threshold effects in virtue of which x has an undesirable tendency. The universal practice of x is a sufficient, but not in general a necessary condition. (It can be a necessary condition in certain cases, if the threshold is high enough; in such cases the general practice—as I have technically defined it—*is* the universal practice of x.) Hence, if there is no general practice of x there are no undesirable threshold

effects, and 'the undesirable consequences that would result from everyone's doing it would be avoided'. Singer's mistake is in thinking that a universal practice is not only a sufficient but also a necessary condition of such effects. Given this misapprehension about threshold effects, it is reasonable for him to claim that less than a universal practice of x will avoid the bad effects. It is reasonable for him to claim that, if *someone* does not do x, then those bad effects will not be produced. This is obviously the condition which Singer wants: that the performance of x simply not be universal. Thus we can understand his clause, 'if not everyone does x', as 'if someone does not do x'.

This general line of interpretation of the significance of GC seems to be borne out also by the passage succeeding the one cited above (p. 66). Singer contrasts GA with a principle with which it might be confused, and emphasizes that the antecedent of GA (and hence, we might add, of GC) concerns 'the *collective consequences* of everyone's acting in a certain way', which he says is 'not always the same' as 'the *individual* consequences of actions of a certain kind'. Let us think of the collective consequences of everyone's acting in a certain way as the tendency, and the individual consequences as the effects of particular acts of the kind having that tendency. (Let us disregard the trivial sense in which Singer's claim is true, that is, the fact that the individual consequences of an act—its effects—are not the same as the consequences of everyone's doing the same—its tendency. For the latter (except when the class has only one member) includes the consequences of several acts.) In saying that these are not the same, Singer may well be saying that there can be a qualitative difference between the tendency and the effects of a given act. Recall that Harrison made this claim and that it was, as we have seen, based upon systematically incomplete descriptions. When we note that Singer does not put a premium upon complete descriptions, but employs instead reasons which are based on incomplete descriptions, we can see that Singer's general approach fits into a pattern of misapprehensions about utilitarian generalization that we have already discussed.

Let us consider now just two of the several difficulties in going from this reconstruction of Singer's approach to the rendering of GC with '. . . someone . . .' in the consequent. Singer's first comment upon his proposed deduction is the following:

> In the above generalization from the principle of consequences, (GC), 'everyone' is treated collectively, not distributively. The hypothesis 'If the consequences of everyone's acting in a certain way would be undesirable' differs from 'If the consequences of *each and every act* of that kind would be undesirable.' (p. 66.)

Here and in the succeeding passage Singer's point is to show (in relation to GC) that an undesirable tendency can imply that some, but not necessarily all, acts of the kind are wrong, and (in relation to GA) that acts can be wrong without having undesirable consequences.

But the passage just cited might suggest that both instances of 'everyone' in GC are to be 'treated collectively'—the instance in the consequent as well as that in the antecedent (or hypothesis). If this is so, there seems to be an equivocation in the deduction. For if the 'everyone' in the consequent of GC were treated collectively, then we could not go from 'not everyone ought to do x' or 'not everyone has the right to do x' to 'someone ought not to do x' or 'someone has not the right to do x'. It would seem that we would be dealing with 'collective' rights, duties, or obligations. For example, 'not everyone has the right to do x' could be understood as denying the existence of a joint right of a collectivity to do x: as 'it is not the case that everyone (together) has the right to do x'.

But this sort of interpretation would stray too far from Singer's obvious intentions. It is clear that, in the case of GC, as necessarily in the case of GP and GA, Singer is dealing with the wrongness of particular acts and not at all with the rights (or duties or obligations) of groups conjointly. Moreover, Singer refers in the passage cited only to the antecedent of GC, and not at all to the consequent. The point of these instructions is apparently to direct our attention to the tendencies of acts and away from the normal or usual effects which acts of a certain kind might have. The comments are simply misleading; but misleadingly inadequate comments are characteristic of this section. For example, Singer fails to give similar instructions for interpreting GA, although clearly the antecedents of GC and GA are identical.

The matter could be clarified somewhat by a partial formalization of the deduction. The 'everyone's operate simply as quantifiers or parts thereof:

GC: If the tendency of x is undesirable, then $(\exists y)$ (it would be wrong for y to do x).
GP: If $(\exists y)$ (it would be wrong for y to do x), then (y) (it would be wrong for y to do x).

GA: If the tendency of x is undesirable, then (y) (it would be wrong for y to do x).

This eliminates the problem about 'everyone'. It also clarifies the 'transition' from 'some' to 'all': it consists in the inference, via GP, from '$(\exists y)(\ldots y\ldots)$' to '$(y)(\ldots y\ldots)$'.

But it must also be admitted that we are so far giving Singer the benefit of the doubt. He is not at all clear about these issues.

The second difficulty I want to mention concerns a more radical transformation of the consequent of GC which is suggested by Singer. GC might be rendered as:

> If the consequences of everyone's doing x would be undesirable, then it ought not to be the case that everyone does x.

How are we to take this? It does not seem to concern particular acts, but general practices as a whole. And it would be impossible to use this in a deduction of GA, since both GP and GA concern the wrongness of particular acts.

It would be implausible to render GP as 'If it ought not to be the case that everyone does x, then everyone ought not to do x (i.e. no one ought to do x)'. For if we take this at face value, assuming that the antecedent is not to be understood as 'If someone ought not to do x', then the antecedent can be true and the consequent false. For example, if everyone fails to prepare a paper for this week's seminar, the consequences will be undesirable; hence it ought not to be the case that everyone fails to prepare a paper for this week's seminar. But from this it does not follow that anyone does a wrong thing in failing to prepare a paper. For papers are assigned by the instructor and prepared by the students. The instructor fails to assign a paper. *He* is wrong, if anyone is. But his wrong act does not consist in failing to prepare a paper; it consists in failing to assign one. Hence GP cannot be understood in this way, for it is not the sort of principle that can be so falsified. (Suggested by G. Nakhnikian.)

If the consequent, 'it ought not to be the case that everyone does x', suggests anything at all, it suggests either that there is some disvalue in everyone's doing x or that the universal practice of x should be prevented. But from neither of these does it follow that anyone's doing x is wrong. It may be that this is the formulation of GC which Singer intended, but it is entirely askew in relation to the whole deduction, in relation to GP and GA, and in relation to C itself, from which it is supposedly derived. This brings us to the principle of consequences, the purported foundation of the deduction.

5. *The principle of consequences, C*

> If the consequences of A's doing x would be undesirable, then it would be wrong for A to do x.

Singer also uses 'A ought not to do x' for the consequent. Interestingly, 'it ought not to be the case that A does x' is not indicated. It

would after all be a very poor rendering of such a simple utilitarian principle, which is used to determine the wrongness of acts on the basis of their negative simple utilities. For if it were rendered in that improbable way, it would suggest, not that *A*'s doing *x* is wrong, but rather that anyone's failure to keep *A* from doing *x* is wrong.

The establishment of GA rests ultimately on C and GP. I have given some account of GP, which seems a perfectly unobjectionable principle. Here is what Singer says of C: that it is

> a necessary ethical or moral principle. It is necessary not only in the sense that its denial involves self-contradiction. It is necessary also in the sense that like the generalization principle, it is a necessary presupposition or precondition of moral reasoning. There can be sensible and fruitful disagreement about matters within the field delimited by it, but there can be no sensible or fruitful disagreement about the principle itself. (p. 64.)

Can we accept such strong claims? It seems perfectly consistent for us to maintain that a particular act is not wrong even though its consequences are undesirable (on the whole). If we aren't strict utilitarians, we can allow that *A*'s reasons for doing *x* are relevant and thus, not just that *A* is not blameworthy, but that his act is not wrong. But even if we are utilitarians, we need not take a negative simple utility as a sufficient condition of an act's being wrong, as is implied by C. For example, the alternatives might have worse consequences. Singer acknowledges this point (with regard to the complement: p. 64; see also pp. 106–7); but it is not reflected in the formulation of C. C is a non-comparative principle; it is also strong. Hence it must yield anomalous judgements in cases where all the alternatives have bad consequences (on the whole). It is impossible to read C in any other way.

Notice that Singer claims that C can be 'misunderstood' (p. 64). The confusions which we might have about C are limited by Singer (pp. 64–65) to the distinction between these two forms:

C_1: If the consequences of *A*'s doing *x* would be undesirable on the whole, then *A* has a conclusive reason against doing *x*.

C_2: If some consequences of *A*'s doing *x* would be undesirable, then *A* has a presumptive (i.e. good or prima facie) reason against doing *x*.

The *ceteris paribus* condition incorporated in C_2 (but not in C_1) rests on the possibility that the accounts may not be reckoned completely.

But note that these are two different principles; they are not alternative formulations; they are not equivalent. One can, for example,

consistently accept C_2 and reject C_1. Is this not sensible disagreement about C? Moreover, if we allow C_1 and C_2, why not also allow:

C_3: If the consequences of *A*'s doing *x* would be undesirable on the whole, then *A* has a presumptive reason against doing *x*.

C_4: If some consequences of *A*'s doing *x* would be undesirable, then *A* has a conclusive reason against doing *x*.

C_3 is the weaker version of C_1 and C_4 is the stronger version of C_2. No two of these four are equivalent. Note how sensible, fruitful, and consistent disagreement about C can be. C_1 is too strong, given its non-comparativeness. C_4 is absurd. C_2 is better, but it seems the manifestation of excessively pessimistic caution; it is unnecessary given C_3. C_3 is the best of the lot, and could readily be accepted by non-utilitarians. But an even better principle for a negative utilitarian such as Singer would be:

(NU) If the consequences of an act would be undesirable (on the whole) and worse than those of some alternative, then it would be wrong for that act to be performed.

And this is clearly not Singer's C, which is C_1 (or C_2? or both?).

In view of the inherent defects of C, GC cannot be an acceptable moral principle if it is in fact derived from C alone; hence GA cannot be an acceptable moral principle either. But let us examine the status of GC more closely.

6. *The status of GC*

We are told that GC is a 'generalization from' C, although it is admittedly not the 'true logical generalization' of C. The latter seems to be:

> If the consequences of each and every act of a certain kind would be undesirable, then each and every act of that kind is wrong.

It is not clear how this is a logical generalization from C. For C is:

> If the consequences of *A*'s doing *x* would be undesirable, then it would be wrong for *A* to do *x*.

If we generalize C we presumably do this:

> For all *A*, for all *x*, if the consequences of *A*'s doing *x* would be undesirable, then it would be wrong for *A* to do *x*.

This involves quantifying over agents and actions. But note that we are quantifying over particular acts, and not over kinds of acts. And yet the 'true logical generalization' of C concerns acts of certain kinds, not just acts. Singer may mean that '*x*' in C is to be a place-holder for

descriptions, but this is not made clear. In any event, I shall show below that the 'true logical generalization' of C is what we get when we apply GP to C.

But what about GC itself? It is hard to see how it can be a 'generalization' from C. If the deduction is to stand, if we are to make a 'transition' from 'some' to 'all', the consequent of GC must be particular, while the antecedent is general. We might however attempt some unorthodox 'generalizing'. Perhaps we should 'generalize' the antecedent and consequent separately. We must go from:

C: If the consequences of *A*'s doing *x* would be undesirable . . .

to:

GC: If the consequences of everyone's doing *x* would be undesirable . . .

which involves simply a change from '*A*' to 'everyone'. Let us do the same with the consequents. The results are very odd.

Begin with these three formulations of the consequent of C given by Singer:

(i) *A* has no right to do *x*;
(ii) *A* ought not to do *x*;
(iii) it would be wrong for *A* to do *x*.

These yield:

(iv) everyone has no right to do *x*;
(v) everyone ought not to do *x*;
(vi) it would be wrong for everyone to do *x*.

(iv)–(vi) are by no means the consequents of GC. (iv) and (v) are, if anything, the consequents of GA! (vi) might also do for GA, but surely not for GC. However, if we shift the negation in (iv) and (v), thus changing their sense, we get two forms of the consequent of GC:

(vii) not everyone has (the) right to do *x*;
(viii) not everyone ought to do *x*.

We need not make such a shift if we use instead, 'it ought not to be the case that —— does *x*', in the blank of which we could insert either '*A*' or 'everyone'. But this is the least plausible rendering of these principles. Is this none the less the operation Singer had in mind? One can only guess. We have a similar difficulty with two other forms which Singer does not exploit: 'it would not be right for —— to do *x*' and 'it is not the case that —— has the right to do *x*'. These are not clearly about particular acts at all.

But it might still be claimed that GC is, in any case, a 'necessary' ethical principle, since the consequent must always be true when the

antecedent is satisfied. If so, the deduction would fare better, whatever we might think, for example, of C. Now why should we say that GC is necessary (or at least true) in the sense that the consequent is true whenever the antecedent is satisfied? We might appeal to GC or to GA. The first begs the question; the second renders the whole deduction circular. We might appeal to C. For if the antecedent of GC is satisfied, undesirable consequences are produced, and at least one act has undesirable consequences on the whole. According to C, that act is wrong. Therefore some x is wrong. This satisfies the consequent of C as we have interpreted it. (Suggested by H.-N. Castañeda.)

The foregoing is a tempting argument, for it provides a place for C in the total deduction of GA. But if we accept such an argument we must also accept C; indeed, we must accept a strong form of C (e.g. C_1). Any such form is unacceptable, however.

There is a more direct, independent argument against GC, which holds against any interpretation of the principle. GC is not the sort of principle we could safely use as a guide to action. For its application—indeed, its successful observance—could be self-defeating. On any interpretation of GC, following the principle requires merely that there not be a universal practice of x, where the tendency of x is undesirable. This condition can be satisfied if we take it that some x is wrong and, e.g., refrain from performing at least one act; or it is satisfied if we prevent the universal practice of x. But even if these conditions were satisfied, the undesirable consequences might still occur. That is, if a general but not universal practice of x is a sufficient condition of undesirable threshold effects, the practice of x could be non-universal and still those effects could occur. The consequent of GC is too weak. It is true that if enough acts x are not performed, those effects will not be produced. But the consequent of GC says nothing about avoiding or preventing or in effect refraining from the general practice of x; it concerns only the universal practice.

In other words, GC is ill-conceived. The rationale underlying GC is presumably the same as that underlying C—as Singer makes clear—to minimize disutilities. We might say that C and GC have a similar point (in the sense in which (AU) and (GU) and perhaps (IRU) do). This is the point of negative utilitarianism. The defect of C is that it is too strong. The defect of GC is that it is too, in another sense, weak. The observance of either would be self-defeating, given a negative utilitarian point of view.

7. *The pointlessness of the deduction*

If one has a system containing a principle like C, there is no need for a principle like GA. (Of course if one wants a convincing negative utilitarian system it would be best to have (NU) or its general utilitarian analogue; but let us concern ourselves with Singer's system as it stands.) For C is extensionally equivalent to the strong version of GA. If we compare C (i.e. C_1) with the weakened GA we find that the only difference in their import is that the class of acts which are counted as wrong by C, all things considered, must be identical with the class of acts counted as prima facie wrong by GA, all things considered. (The weakened GA is in fact extensionally equivalent to C_3.)

This is not the sort of criticism which is usually made against a system which includes two such principles. The expected criticism is that C and GA conflict; therefore the system is inconsistent. In fact, since GA is weak, its implications could not be incompatible with those of any other principle. But that is not the main reason why Singer's principles are compatible. Nor is it simply that C and GA have compatible but overlapping implications. For example, it might be supposed—as Singer suggests (pp. 105–6)—that every act which C counts as wrong GA counts as wrong, but that GA counts more acts wrong, including some without undesirable effects. But this is an illusion fostered by the method of rebuttals. Reasons can be generated against a class of acts which C cannot condemn, but these reasons are necessarily rebuttable. That is to say, when the acts are fully described, if GA counts them wrong then C does also. And if we simply want to generate reasons against acts based on some of their relevant features, we could do a similar thing by employing C_2 instead of C_1.

Consequently, GA is superfluous in the system. And, since it is employed in accordance with a method of rebuttals which is liable to be grossly misleading, its adoption and application would seem to yield more disadvantage than not.

8. *Transforming the deduction*

We should not expect GP to produce anything novel or interesting when combined with a given principle. It can add nothing of substance to the import of C or GC, for it has not the same sort of import. Moreover, its employment is already presupposed when we apply any such principle or argument. If *A*'s doing *x* would be wrong, then any relevantly similar act (performed in relevantly similar circumstances by a relevantly similar agent) would be wrong. This can be taken in two ways. Let the act involved in *A*'s doing *x* be *z*. Now *z* is wrong because it has bad effects. One who holds to C should

consistently hold that any such act, i.e. an act with bad effects on the whole, is wrong. This is one way in which he treats like cases alike. But since z is counted as wrong because of its bad effects, it has some collection of simple (or general) utilitarian properties Z in virtue of which it has such effects. The second way to treat like cases alike is to count any act Z as wrong—unless of course some instance of Z is relevantly different and therefore does not have bad effects on the whole, in which case that instance of Z must be treated differently.

This does not yield a form of utilitarian generalization, but what Singer calls the 'true logical generalization' of C:

> If the consequences of each and every act of a certain kind (Z) would be undesirable, then each and every act of that kind is wrong.

Now what is the difference between the foregoing principle and the strong form of GA? The difference seems to be (on a straightforward utilitarian account) merely an immaterial and perhaps misleading difference in formulation. For GA cannot condemn acts which do not have undesirable effects on the whole, while the foregoing principle can also condemn acts because of their threshold-related disutilities and thus can account for threshold effects.

This is the closest I can come to developing a 'deduction' of GA from C. It should also be observed, however, that a deduction along the lines proposed by Singer is possible, employing a premiss slightly different from GC:

TH: If the consequences of everyone's doing x would be undesirable, then it would be wrong for anyone whose performance of x would contribute to the production of those undesirable consequences to do x.

[AP: If it is wrong for anyone, whose performance of x would contribute to the production of the undesirable consequences produced by everyone's doing x, to do x, then it would be wrong for someone to do x.]

GP: If it would be wrong for someone to do x, then it would be wrong for anyone to do x.

GA: If the consequences of everyone's doing x would be undesirable, then it would be wrong for anyone to do x.

TH does not suffer the main defect of GC, namely, that its consequent is too weak. TH does what GC fails to do: it provides for thresholds. The extra (bracketed) step, AP, is a formal premiss linking the consequent of TH with the antecedent of GP. The consequent of TH is actually stronger than the antecedent of GP: it indicates not merely

that someone should not do x, but that no acts having threshold-related disutilities should be performed. But if and whenever the consequent of TH is satisfied, then necessarily the antecedent of GP is satisfied.

TH should be extensionally equivalent to C. For its import is that acts with negative simple utilities are wrong—whether or not threshold effects are involved. Actually, the consequent of TH may be too strong; and if so, there would not be this extensional equivalence. It might be necessary to amend TH so that it condemns only those acts whose undesirable effects outweigh their desirable ones. But on the other hand, if we place a premium upon completeness of descriptions, we can prevent there being such differences among instances of x that some acts x do not have undesirable consequences on the whole even though they may be said to contribute to the relevant undesirable threshold effects. And if we set this condition upon acceptable descriptions, we need no *ceteris paribus* qualifier. The result is a strong GA, which is extensionally equivalent to both TH and C.

The moral is: GP adds nothing whatever to the deduction.

BIBLIOGRAPHY

A. *Works Cited in the Text*

THE author wishes to express his gratitude to the writers and publishers listed below who have given their kind permission to reprint passages quoted in the text.

BAIER, KURT. *The Moral Point of View: A Rational Basis of Ethics.* Ithaca, New York: Cornell University Press, 1958.

BRANDT, RICHARD B. *Ethical Theory: The Problems of Normative and Critical Ethics.* Englewood Cliffs, New Jersey: Prentice-Hall, 1959. See pp. 396 ff.

—— 'Toward a Credible Form of Utilitarianism', in *Morality and the Language of Conduct*, ed. by H.-N. Castañeda and G. Nakhnikian. Detroit: Wayne State University Press, 1963, pp. 107–44. (Passage reprinted by permission of The Wayne State University Press and the author. Copyright © 1963 by Wayne State University Press, Detroit 2, Michigan. All rights reserved.)

BROAD, C. D. 'On the Function of False Hypotheses in Ethics', *International Journal of Ethics*, xxvi (April 1916), 377–97. (Passages reprinted by permission of The University of Chicago Press, publisher. Copyright by The University of Chicago.)

FEIGL, H. 'Validation and Vindication', in *Readings in Ethical Theory*, ed. by W. Sellars and J. Hospers. New York: Appleton-Century-Crofts, 1952, pp. 667–80.

HARRISON, JONATHAN. 'Utilitarianism, Universalisation, and Our Duty to Be Just', *Proceedings of the Aristotelian Society*, liii (1952–3), 105–34. (Passages reprinted by courtesy of the Editor of The Aristotelian Society and the author.)

HARROD, R. F. 'Utilitarianism Revised', *Mind*, xlv (April 1936), 137–56. (Passage reprinted by permission of the Editor of *Mind* and the author.)

HART, H. L. A. 'Are There Any Natural Rights?', *Philosophical Review*, lxiv (April 1955), 175–91. (Passage reprinted by permission of the *Philosophical Review*.)

MOORE, G. E. *Ethics.* London: Oxford University Press (many editions).

—— *Principia Ethica.* Cambridge: Cambridge University Press (many editions).

RAWLS, JOHN. 'Justice as Fairness', *Philosophical Review*, lxvii (April 1958), 164–94. (Passages reprinted by permission of the *Philosophical Review*.)

—— 'Two Concepts of Rules', *Philosophical Review*, lxiv (January 1955), 3–32. (Passages reprinted by permission of the *Philosophical Review*.)

SINGER, MARCUS G. *Generalization in Ethics: An Essay in the Logic of Ethics, with the Rudiments of a System of Moral Philosophy.* New York: Alfred A. Knopf, 1961. (Passages reprinted with permission of Alfred A. Knopf, Inc. Copyright © 1961 by Marcus G. Singer.)

SMART, J. J. C. 'Extreme and Restricted Utilitarianism', *Philosophical Quarterly*, vi (October 1956), 344–54. (Passages reprinted by permission of the Editor of the *Philosophical Quarterly*.)

STOUT, A. K. 'But Suppose Everyone Did the Same', *Australasian Journal of Philosophy*, xxxii (May 1954), 1–29. (Passage reprinted by permission of the Editor of the *Australasian Journal of Philosophy* and the author.)

STRANG, COLIN. 'What if Everyone Did That?', *Durham University Journal*, xxiii (N.S., December 1960), 5–10.

TAYLOR, PAUL W. *Normative Discourse*. Englewood Cliffs, New Jersey: Prentice-Hall, 1961.

TOULMIN, STEPHEN E. *An Examination of the Place of Reason in Ethics*. Cambridge: Cambridge University Press, 1950.

B. *Relevant Works Not Cited in the Text (Brief List)*

AUSTIN, JOHN. *The Province of Jurisprudence Determined* and *The Uses of the Study of Jurisprudence*. London: Weidenfeld and Nicolson, 1954. Austin was, I believe, the first to use the term 'tendency' in the way suggested above. Austin's use is probabilistic, however. The tendency of an act is, for him, 'the sum of its probable consequences, in so far as they are important and material'. See Lecture ii, pp. 38–40, 47–48.

DIGGS, B. J. 'Rules and Utilitarianism', *American Philosophical Quarterly*, i (January 1964), 32–44.

EWING, A. C. 'What Would Happen if Everybody Acted Like Me?', *Philosophy*, xxviii (January 1953), 16–29.

HOSPERS, JOHN. *Human Conduct: An Introduction to the Problems of Ethics*. New York: Harcourt, Brace and World, 1961. See sec. 17.

MABBOTT, J. D. 'Moral Rules', *Proceedings of the British Academy*, xxxix (1953), 97–118.

MCCLOSKEY, H. J. 'An Examination of Restricted Utilitarianism', *Philosophical Review*, lxvi (October 1957), 466–85.

NAKHNIKIAN, GEORGE. 'Generalization in Ethics', *Review of Metaphysics*, xvii (March 1964), 436–61. The most extensive review of Singer's book.

NOWELL-SMITH, P. H. *Ethics*. London: Penguin Books, 1954. See ch. 16.

SINGER, MARCUS G. 'Generalization in Ethics', *Mind*, lxiv (July 1955), 361–75.

URMSON, J. O. 'The Interpretation of the Philosophy of J. S. Mill', *Philosophical Quarterly*, iii (January 1953), 33–39. A rule-utilitarian reading of Mill.

WASSERSTROM, RICHARD A. *The Judicial Decision: Toward a Theory of Legal Justification*. Stanford: Stanford University Press, 1961. Argues for the equivalence of the two kinds of utilitarianism in the realm of judicial decision.

INDEX

Page references in **bold type** indicate places where expressions are defined or explained.


9 780198 241973