





# UTILITARIANISM

## for and against

J. J. C. SMART

*Emeritus Professor, University of Adelaide*  
*Professor of Philosophy, Australian National University*

BERNARD WILLIAMS

*Provost of King's College, Cambridge*



**CAMBRIDGE**  
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS  
Cambridge, New York, Melbourne, Madrid, Cape Town,  
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press  
32 Avenue of the Americas, New York NY 10013-2473, USA

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521098229](http://www.cambridge.org/9780521098229)

© Cambridge University Press 1973

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1973  
27th printing 2008

*A catalogue record for this publication is available from the British Library.*

ISBN 978-0-521-09822-9 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

## Contents

### **An outline of a system of utilitarian ethics**

J. J. C. SMART

1	Introductory	3
2	Act-utilitarianism and rule-utilitarianism	9
3	Hedonistic and non-hedonistic utilitarianism	12
4	Average happiness versus total happiness	27
5	Negative utilitarianism	28
6	Rightness and wrongness of actions	30
7	The place of rules in act-utilitarianism	42
8	Simple application of game-theory technique	57
9	Utilitarianism and the future	62
10	Utilitarianism and justice	67

### **A critique of utilitarianism**

BERNARD WILLIAMS

1	Introductory	77
2	The structure of consequentialism	82
3	Negative responsibility: and two examples	93
4	Two kinds of remoter effect	100
5	Integrity	108
6	The indirect pursuit of utility	118
7	Social choice	135

	Bibliography	151
--	--------------	-----

J. J. C. SMART



**An outline of a system of  
utilitarian ethics**

J. J. C. SMART



## I. Introductory

Such writers as J. S. Mill, H. Sidgwick and G. E. Moore, as a result of philosophical reflection, produced systems of normative ethics. Of recent years normative ethics has become distinguished from meta-ethics, which discusses the nature of ethical concepts. Indeed, as a result of the prevalence of 'non-cognitivist' theories of meta-ethics, for example those of C. L. Stevenson<sup>1</sup> and R. M. Hare,<sup>2</sup> normative ethics has fallen into some disrepute, at any rate as a philosophical discipline. For non-cognitivist theories of ethics imply that our ultimate ethical principles depend on our ultimate attitudes and preferences. Ultimate ethical principles therefore seem to lie within the fields of personal decision, persuasion, advice and propaganda, but not within the field of academic philosophy.

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

<sup>1</sup> *Ethics and Language* (Yale University Press, New Haven, 1944).

<sup>2</sup> *The Language of Morals* (Oxford University Press, London, 1952).

common, and often specious, objections to them can be avoided.

It will be my object in the present study to state a system of ethics which is free from traditional and theological associations. This is that type of utilitarianism which R. B. Brandt has called 'act-utilitarianism'.<sup>1</sup> Roughly speaking, act-utilitarianism is the view that the rightness or wrongness of an action depends only on the total goodness or badness of its consequences, i.e. on the effect of the action on the welfare of all human beings (or perhaps all sentient beings). The best sustained exposition of act-utilitarianism is, I think, that in Sidgwick's *Methods of Ethics*,<sup>2</sup> but Sidgwick stated it within the framework of a cognitivist meta-ethics which supposed that the ultimate act-utilitarian principles could be known to be true by some sort of intellectual intuition. I reject Sidgwick's meta-ethics for familiar reasons, and for the purpose of this study will assume the truth of some such 'non-cognitivist' meta-ethical analysis as that of Hare's *Language of Morals*, or possibly that of D. H. Monro in his *Empiricism and Ethics*.<sup>3</sup> (Monro's theory should perhaps be classed as subjectivist rather than as non-cognitivist. However I am inclined to think that in the present state of linguistic theory it is not possible to make a very sharp distinction between these two sorts of theory.<sup>4</sup> For our present purposes the distinction is unimportant, because both sorts of theory imply that a man's ultimate ethical principles depend on his attitudes or feelings.) In adopting such a meta-ethics, I do, of course, renounce the attempt to

<sup>1</sup> See R. B. Brandt, *Ethical Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1959), p. 380. Brandt distinguishes 'act' utilitarianism from 'rule' utilitarianism.

<sup>2</sup> H. Sidgwick, *Methods of Ethics*, 7th ed. (Macmillan, London, 1962).

<sup>3</sup> D. H. Monro, *Empiricism and Ethics* (Cambridge University Press, London, 1967).

<sup>4</sup> See my review of Monro's book, *Philosophical Review* 78 (1969) 259-61.

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

In the first place, B. H. Medlin<sup>1</sup> has argued that it is impossible to state ethical egoism without either confusion or else a sort of pragmatic inconsistency. Secondly, some widespread ethical systems depend partly on metaphysical premisses, and can therefore be undermined by philosophical criticism of these metaphysical bases. I myself would be prepared to argue that this is the case with respect to so-called 'natural law' ethics, which depends on a quasi-Aristotelian metaphysics. Thirdly, any system of deontological ethics, that is any system which does not appeal to the consequences of our actions, but which appeals to conformity with certain rules of duty, is open to a persuasive type of objection which may well be found convincing by some of those people who have the welfare of humanity at heart. For though, conceivably, in most cases the dictates of a deontological ethics might coincide with those of human welfare and of an act-utilitarian ethics, there must be some possible cases in which the dictates of the system clash with those of human welfare, indeed in which the deontological principles prescribe actions which lead to avoidable human misery. In the most attractive forms of deontological ethics the conflict with utilitarianism is in consequence of some principle of 'justice' or 'fairness', and I shall revert to this issue later.<sup>2</sup> In other cases,

<sup>1</sup> 'Ultimate principles and ethical egoism', *Australasian Journal of Philosophy* 35 (1957) 111-18.

<sup>2</sup> See pp. 67-73 below.

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

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

<sup>1</sup> Sir David Ross, *Foundations of Ethics* (Oxford University Press, London, 1939).

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

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

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

<sup>1</sup> See pp. 72-3 below.

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

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

<sup>1</sup> 'A note on act utilitarianism', *Philosophical Review* 73 (1964) 243-7.

<sup>2</sup> *The Language of Morals*.

<sup>3</sup> See R. Montague, 'Logical necessity, physical necessity, ethics, and quantifiers', *Inquiry* 3 (1960) 259-69.

may not really be worse off than the cognitivist. At any rate, I am assuming in this monograph that adequate non-cognitivist theories of meta-ethics exist.

## 2. Act-utilitarianism and rule-utilitarianism

The system of normative ethics which I am here concerned to defend is, as I have said earlier, *act-utilitarianism*. Act-utilitarianism is to be contrasted with rule-utilitarianism. Act-utilitarianism is the view that the rightness or wrongness of an action is to be judged by the consequences, good or bad, of the action itself. Rule-utilitarianism is the view that the rightness or wrongness of an action is to be judged by the goodness and badness of the consequences of a rule that everyone should perform the action in like circumstances. There are two sub-varieties of rule-utilitarianism according to whether one construes 'rule' here as 'actual rule' or 'possible rule'. With the former, one gets a view like that of S. E. Toulmin<sup>1</sup> and with the latter, one like Kant's.<sup>2</sup> That is, if it is permissible to interpret Kant's principle 'Act only on that maxim through which you can at the same time will that it should become a universal law' as 'Act only on that maxim which you as a humane and benevolent person would like to see established as a universal law.' Of course Kant would resist this appeal to human feeling, but it seems necessary in order to interpret his doctrine in a plausible way. A subtle version of the Kantian type of rule-utilitarianism is given by R. F. Harrod in his 'Utilitarianism Revised'.<sup>3</sup>

<sup>1</sup> *An Examination of the Place of Reason in Ethics* (Cambridge University Press, London, 1950).

<sup>2</sup> Immanuel Kant, *Groundwork of the Metaphysics of Morals*. Translated from the German in *The Moral Law*, by H. J. Paton (Hutchinson, London, 1948).

<sup>3</sup> *Mind* 45 (1936) 137-56.

I have argued elsewhere<sup>1</sup> the objections to rule-utilitarianism as compared with act-utilitarianism.<sup>2</sup> Briefly they boil down to the accusation of rule worship:<sup>3</sup> the rule-utilitarian presumably advocates his principle because he is ultimately concerned with human happiness: why then should he advocate abiding by a rule when he knows that it will not in the present case be most beneficial to abide by it? The reply that in most cases it is most beneficial to abide by the rule seems irrelevant. And so is the reply that it would be better that everybody should abide by the rule than that nobody should. This is to suppose that the only alternative to 'everybody does *A*' is 'no one does *A*'. But clearly we have the possibility 'some people do *A* and some don't'. Hence to refuse to break a generally beneficial rule in those cases in which it is not most beneficial to obey it seems irrational and to be a case of rule worship.

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

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

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

<sup>2</sup> For another discussion of what in effect is the same problem see A. K. Stout's excellent paper, 'But suppose everyone did the same', *Australasian Journal of Philosophy* 32 (1954) 1-29.

<sup>3</sup> On rule worship see I. M. Crombie, 'Social clockwork and utilitarian morality', in D. M. Mackinnon (ed.), *Christian Faith and Communist Faith* (Macmillan, London, 1953). See p. 109.

<sup>4</sup> David Lyons, *The Forms and Limits of Utilitarianism* (Oxford University Press, London, 1965). Rather similar considerations have been put

that an exception to a rule  $R$  produces the best possible consequences. Then this is evidence that the rule  $R$  should be modified so as to allow this exception. Thus we get a new rule of the form 'do  $R$  except in circumstances of the sort  $C$ '. That is, whatever would lead the act-utilitarian to break a rule would lead the Kantian rule-utilitarian to modify the rule. Thus an adequate rule-utilitarianism would be extensionally equivalent to act-utilitarianism.

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

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

---

forward by R. M. Hare, *Freedom and Reason* (Oxford University Press, London, 1963), pp. 131-6, and R. B. Brandt, 'Toward a credible form of utilitarianism', in H. N. Castañeda and G. Nakhnikian, *Morality and the Language of Conduct* (Wayne State University Press, Detroit, 1963), esp. pp. 119-23.